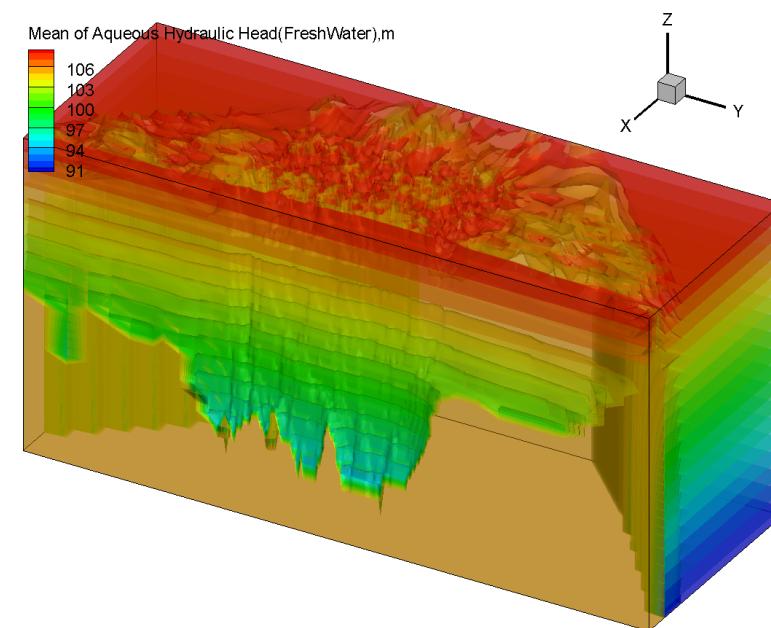
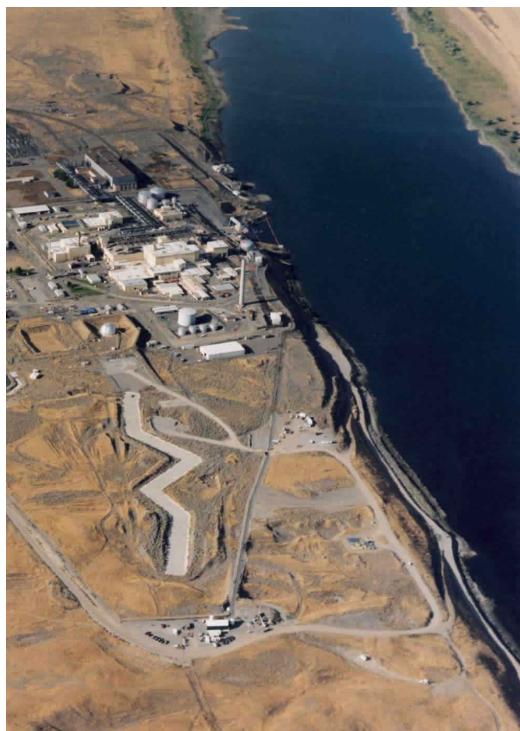


# Uncertainty Quantification Algorithms, Analysis and Applications for High Dimensional Stochastic PDE Systems

Guang Lin, Pacific Northwest National Laboratory



# Uncertainty versus Error

**Uncertainty:** A potential deficiency in any phase or activity of a modeling process that is due to lack of knowledge.

Aleatory (irreducible) Uncertainty  
Epistemic (reducible) Uncertainty

**Error:** A recognizable deficiency in any phase or activity of modeling and simulation that is **not** due to lack of knowledge.

**Sources of Uncertainty:** Initial and Boundary Conditions, Thermo-physical/Structural Properties, Geometric Roughness, Interaction Forces, Background Noise, ...



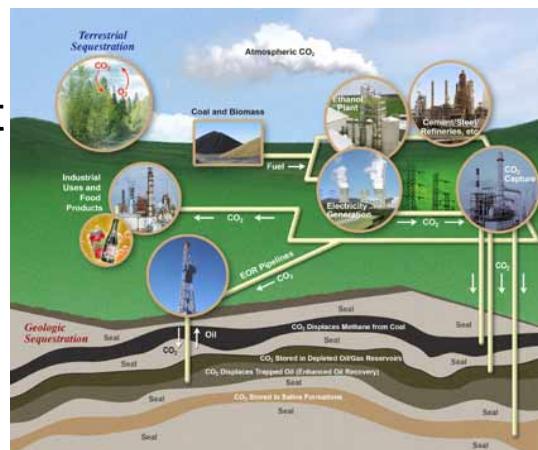
# Objective of Uncertainty Quantification

- ▶ Introduce error bars into numerical simulations.
- ▶ Understand the propagation of uncertainty in a dynamical system.
- ▶ Assessment of the stochastic response.
  - Desired statistics.
  - Reliability analysis.
  - Sensitivity analysis.
  - .....

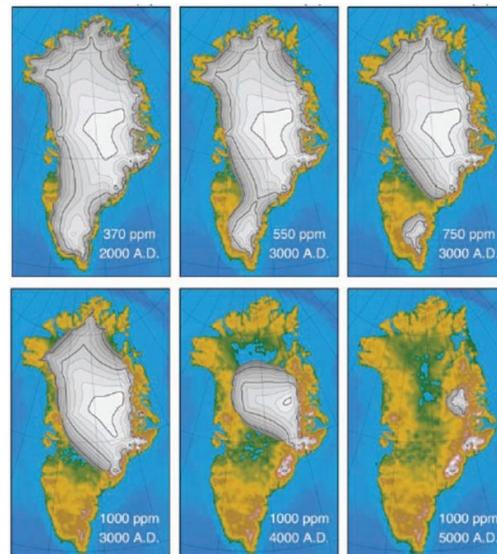
# Uncertainty Quantification for Predictive Modeling of Complex Systems

- ▶ Vision - Transform our ability to uncertainty quantification, model verification, validation and calibration of complex systems
- ▶ Outcomes - Provide fundamental understanding to enable

Safe and efficient remediation and CO<sub>2</sub> sequestration strategies



Better understand the ice sheet dynamics and its interaction with climate



Better prediction of climate changes



Better prediction and control of power system stability and reliability



# Outline:

- ❖ Uncertainty Quantification Algorithms – generalized Polynomial Chaos
- ❖ Open Issues of Uncertainty Quantification Algorithms
- ❖ Subsurface Uncertainty Quantification Application
- ❖ Algorithms and Climate Application for Optimal Parameter Estimation (Model Calibration)
- ❖ Summary

## Generalized Polynomial Chaos - gPC

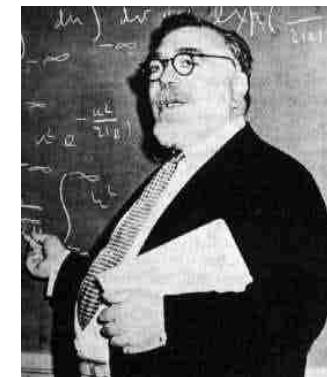
$$T(x, t; \omega) = \sum_{j=0}^{\infty} T_j(x, t) \Phi_j(\xi(\omega))$$

- Polynomials of random variable  $\xi(\omega)$
- Orthogonality :  $\langle \Phi_i \Phi_j \rangle = \langle \Phi_i^2 \rangle \delta_{ij}$

$$\langle f(\xi)g(\xi) \rangle = \int f(\xi)g(\xi)W(\xi)d\xi$$

$$\langle f(\xi)g(\xi) \rangle = \sum_i f(\xi_i)g(\xi_i)w(\xi_i)$$

- Weight function determines underlying random variable (*not necessarily Gaussian*)
- Complete basis from *Askey scheme*
- Each set of basis converges in  $L^2$  sense



Continuous Cases

| Polynomials | Distribution of |
|-------------|-----------------|
| Hermite     | Gaussian        |
| Laguerre    | Gamma           |
| Jacobi      | Beta            |
| Legendre    | Uniform         |

Discrete Cases

| Polynomials | Distribution of   |
|-------------|-------------------|
| Charlier    | Poisson           |
| Krawtchouk  | Binomial          |
| Meixner     | Negative binomial |
| Hahn        | Hypergeometric    |

Xiu & Karniadakis SIAM J. Sci. Comput. 24(2) (2002)

# Implementation of gPC method

$$L(x, u; \xi) = f(x)$$

## ➤ Galerkin Projection:

- PC expansion:  $u = \sum_{|\alpha|=0}^p u_\alpha \phi_\alpha$
- Residual:  $R(\xi) = L(x, \sum_{|\alpha|=0}^p u_\alpha \phi_\alpha) - f(x)$
- Deterministic system of  $u_\alpha$ :  $\boxed{E[R(\xi)\phi_\beta(\xi)] = 0, |\beta| \leq p}$

## ➤ Collocation Projection:

- Interpolation operator:  $\{\xi^{(i)}\}_{i=1}^{Ng}$  a set of grid points in parameter space.
- Deterministic system on grid points:  $\boxed{L(x, u; \xi^{(i)}) = f(x)}$
- Choices of grid points: *full tensor products of Gauss quadrature points –  $O(N^M)$* ; *sparse grids –  $O(N \log(N)^{M-1})$*

# Computational Speed-Up

Lucor & Karniadakis, Generalized Polynomial Chaos and Random Oscillators  
*Int. J. Num. Meth. Eng., vol. 60, 2004*

| PDF      | Error<br>(mean) | Monte-<br>Carlo: M | GPC:<br>(P+1) | Speed-Up   |
|----------|-----------------|--------------------|---------------|------------|
| Gaussian | 2%              | 350                | 56            | 6.25       |
|          | 0.8%            | 2,150              | 120           | 18         |
|          | 0.2%            | 33,200             | 220           | 151        |
| Uniform  | 0.2%            | 13,000             | 10            | 1,300      |
|          | 0.018%          | 1,580,000          | 20            | 79,000     |
|          | 0.001%          | 610,000,000        | 35            | 17,430,000 |



Pacific Northwest  
NATIONAL LABORATORY

# Advantage of gPC

- Fast convergence due to spectral expansion.
- Efficiency due to orthogonality.

$$\begin{aligned}\frac{\partial u}{\partial t} + (u \cdot \nabla) u &= -\nabla p + \nu(1 + \delta\xi)\nabla^2 u \\ \nabla \cdot u &= 0\end{aligned}$$

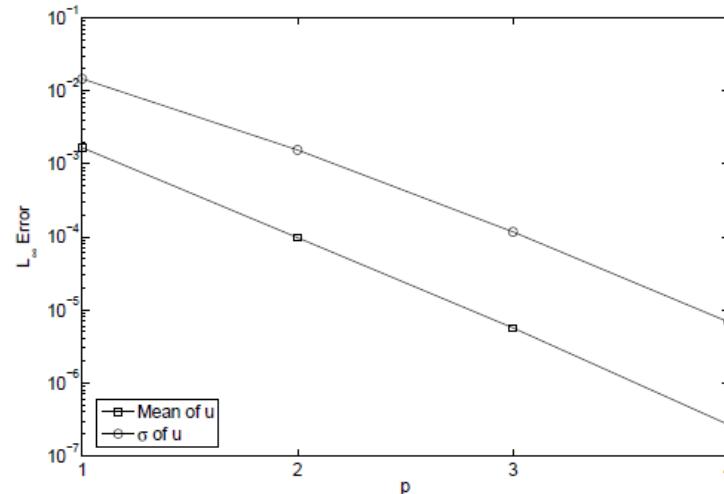
Kovasznay Flow:

$$u = 1 - e^{\lambda x} \cos 2\pi y$$

$$v = \frac{\lambda}{2\pi} e^{\lambda x} \sin 2\pi x$$

$$\lambda = \frac{Re(\xi)}{2} - \left(\frac{Re^2(\xi)}{4} + 4\pi\right)^{1/2}$$

$\xi$  : random variable of Beta(1,1).



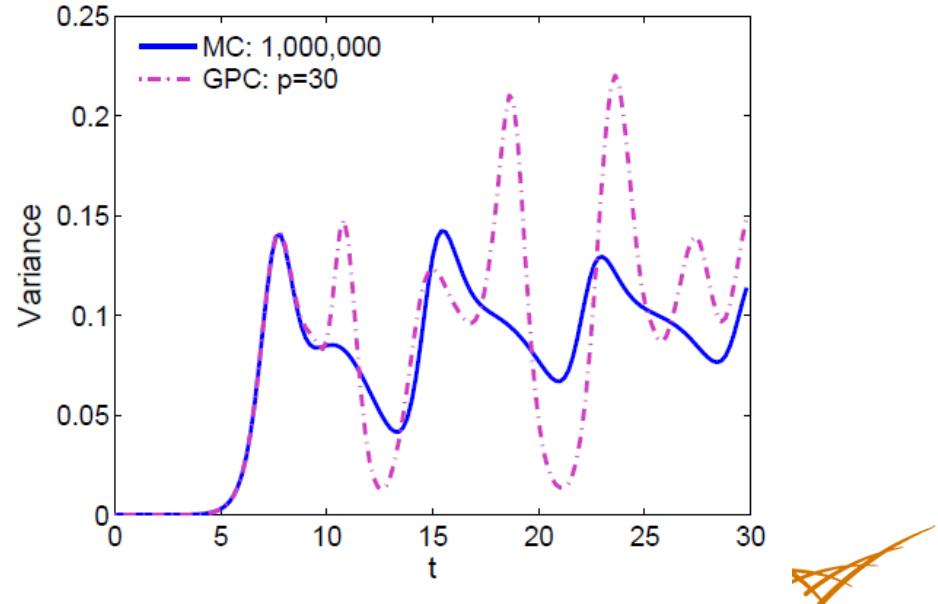
Pacific Northwest  
NATIONAL LABORATORY

# Limitations of gPC

- :( Inefficient for problems with low regularity in the parametric space.
- :( May diverge for long-time integrations.

Kraichnan-Orszag three-mode model:

$$\left\{ \begin{array}{l} \frac{dY_1}{dt} = Y_2 Y_3 \\ \frac{dY_2}{dt} = Y_1 Y_3 \\ \frac{dY_3}{dt} = -2Y_2 Y_3 \end{array} \right. \quad \text{random initial conditions.}$$



# Open Issues

## 1 Parametric Discontinuities/Bifurcations

- ❖ Multi-Element generalized Polynomial Chaos (ME-gPC)
- ❖ Multi-Element PCM (ME-PCM)

## 2 High-Dimensions

- ❖ Sparse Grids
- ❖ ME-PCM with ANOVA
- ❖ Sensitivity Analysis based UQ method

## 3 Scalability on Exascale Platforms

- ❖ Multilevel Parallelism for UQ at the Exascale
- ❖ Stochastic Mutigrid Method

## 4 Uncertainty Quantification in Random Composites

- ❖ Combine random domain decomposition with multi-element probabilistic collocation method

## Open Issue 1: Parametric Discontinuities/Bifurcations

- Multi-Element Probabilistic Collocation Method  
(ME-PCM)

- Decompose  $\Gamma$  into non-overlapping elements  $B^i$

- Define  $A_k = \mathbf{Y}^{-1}(B^k)$

- Define new random variable  $\eta_k : A_k \rightarrow B_k$  on the restricted space

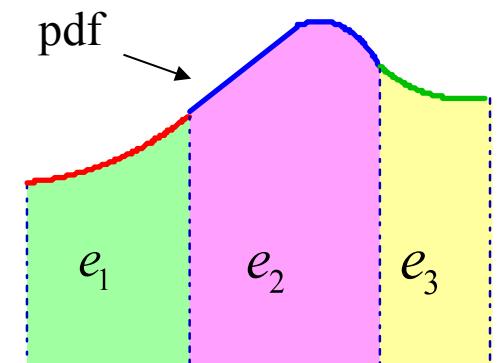
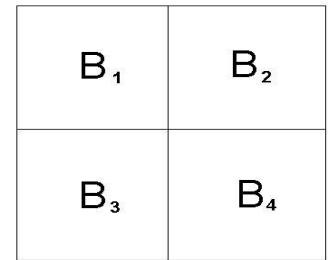
$(A_k, \mathcal{F} \cap A_k, P(\cdot | A_k))$  with conditional PDF  $\hat{\rho}(x | A_k) = \frac{\rho(x)}{P(A_k)}$

- Numerically reconstruct local polynomial chaos basis on each element, orthogonal with respect to  $\hat{\rho}$

- Perform PCM on each element. No  $C^0$  requirement on boundaries (measure 0).

$$\tilde{u}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N_e} \mathcal{I}_{B^i} u_k(\mathbf{x}, \mathbf{y}) \mathbb{I}_{\{\mathbf{y} \in B^i\}} \quad \forall \mathbf{x} \in \overline{D}, \forall \mathbf{y} \in \Gamma$$

choose spatial discretization method



## Moments convergence rate of ME-PCM

**Theorem** Suppose  $f \in \mathcal{W}^{m+1,\infty}(\Gamma)$  and let  $\{B^i\}_{i=1}^{N_e}$  be a nonoverlapping mesh of  $\Gamma$ . Let  $h$  indicate the maximum side length of each element and  $\mathcal{Q}_m^\Gamma$  a quadrature rule with degree of exactness  $m$  in domain  $\Gamma$ . The following error estimate holds:

$$\left| \int_{\Gamma} f(\mathbf{x}) d\mathbf{x} - \sum_{i=1}^{N_e} \mathcal{Q}_m^{B^i} f \right| \leq Ch^{m+1} \|E_\Gamma\|_{m+1,\infty,\Gamma} |f|_{m+1,\infty,\Gamma}$$

where  $C$  is a constant and  $\|E_\Gamma\|_{m+1,\infty,\Gamma}$  denotes the norm of the error functional of  $\mathcal{Q}_m^\Gamma$

$$E_A(g) \equiv \int_A g(\mathbf{x}) d\mathbf{x} - \mathcal{Q}_n^A g, \text{ and } \|E_A\|_{k,\infty,A} = \sup_{\|g\|_{k,\infty,A} \leq 1} |E_A(g)|$$

# Open Issues

## 1 Parametric Discontinuities/Bifurcations

- ❖ Multi-Element generalized Polynomial Chaos (ME-gPC)
- ❖ Multi-Element PCM (ME-PCM)

## 2 High-Dimensions

- ❖ Sparse Grids
- ❖ ME-PCM with ANOVA
- ❖ Sensitivity Analysis based UQ method

## 3 Scalability on Exascale Platforms

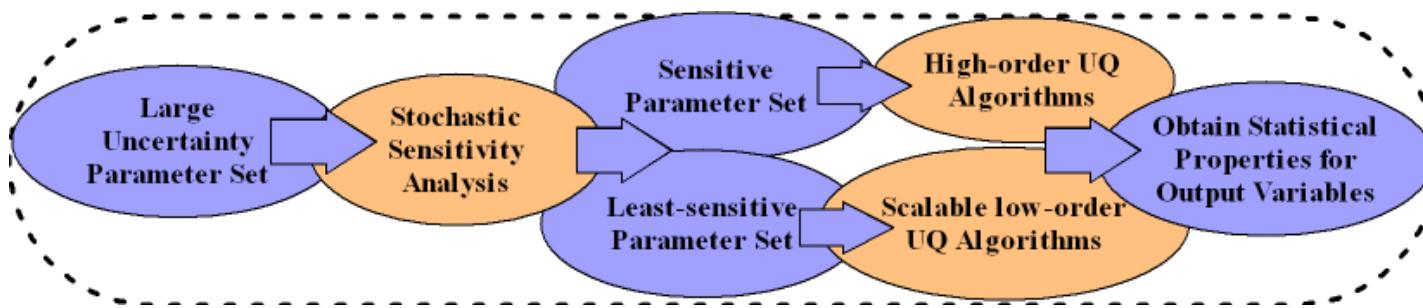
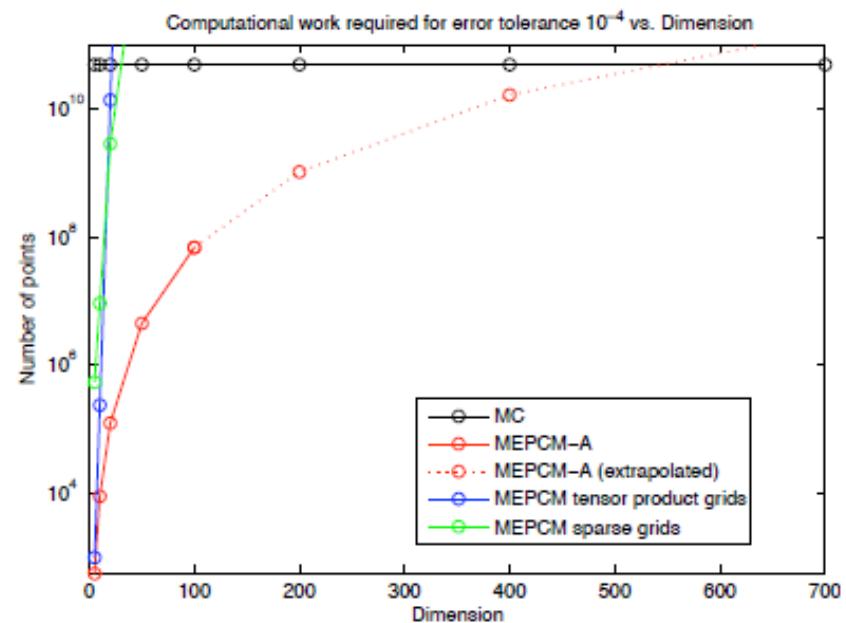
- ❖ Multilevel Parallelism for UQ at the Exascale
- ❖ Stochastic Mutigrid Method

## 4 Uncertainty Quantification in Random Composites

- ❖ Combine random domain decomposition with multi-element probabilistic collocation method

## Open Issue 2: Curse of Dimensionality – UQ Algorithms Development

- ▶ Sparse Grids for Moderate High Dimensional Problem
- ▶ Combine Multi-Element Probabilistic Collocation Method with ANOVA for High Dimensional Problem
- ▶ Sensitivity-Based UQ Method for High Dimensional Problem
  - Most-Sensitivity Parameter Set: High-Order UQ Algorithms
  - Least-Sensitivity Parameter Set: Scalable UQ Algorithms



## Open Issue 2: Curse of Dimensionality – Multi-Element Probabilistic Collocation Method (ME-PCM)

Choice of N-dimensional approximation operator:  $\mathcal{I}_{B^i}$

- Tensor product Lagrangian interpolation

interpolation orders

$$L_{B^i}^P u_k(\mathbf{x}, \mathbf{y}) = \sum_{m=1}^r u_k(\mathbf{x}, \mathbf{q}_m) \cdot l_m(\mathbf{y})$$

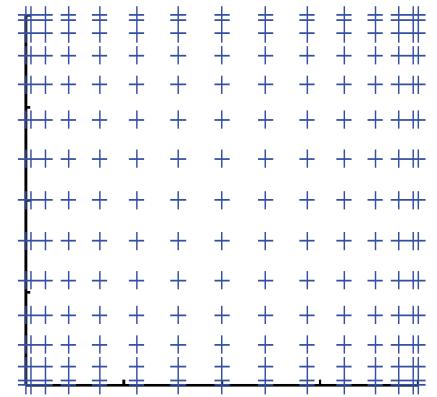
↑  
interpolation points

- Smolyak sparse grid approximation (Smolyak, 1963)

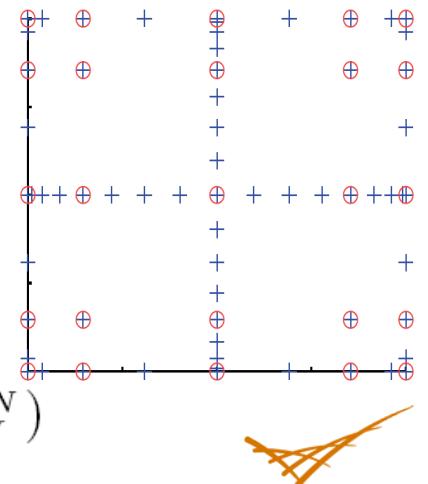
$$\mathcal{V}_j^i(v) = \sum_{m=1}^{n_i} v(y_m^i) \cdot a_m^i \quad \text{1D interp. rule in dimension i}$$

$$\mathcal{S}_{B^i}(s) = \sum_{s-N+1 \leq |\mathbf{i}| \leq s} (-1)^{s-|\mathbf{i}|} \binom{N-1}{s-|\mathbf{i}|} \cdot (\mathcal{V}_1^{i_1} \otimes \dots \otimes \mathcal{V}_N^{i_N})$$

tensor product



sparse



Pacific Northwest  
NATIONAL LABORATORY

interpolatory for nested 1D rules

## Open Issue 2: Curse of Dimensionality – Extension: MEPCM-A for higher-dimensional problems

Using ANOVA-type decompositions: based on the hierarchical decomposition of a multidimensional function into combinations of functions of subgroups of its dimensions (Hoeffding 1948):

$$\mathcal{I}_\mu f(x_1, x_2, \dots, x_N) =$$

$$\begin{aligned} \mathcal{I}_\mu f_0 + \sum_{j_1, (\nu=1)}^N \mathcal{I}_\mu f_{j_1}(x_{j_1}) + \sum_{j_1 < j_2, (\nu=2)}^N \mathcal{I}_\mu f_{j_1, j_2}(x_{j_1}, x_{j_2}) + \\ \sum_{j_1 < j_2 < j_3, (\nu=3)}^N \mathcal{I}_\mu f_{j_1, j_2, j_3}(x_{j_1}, x_{j_2}, x_{j_3}) + \dots \mathcal{I}_\mu f_{j_1, \dots, j_N}(x_{j_1}, \dots, x_{j_N}) \end{aligned}$$

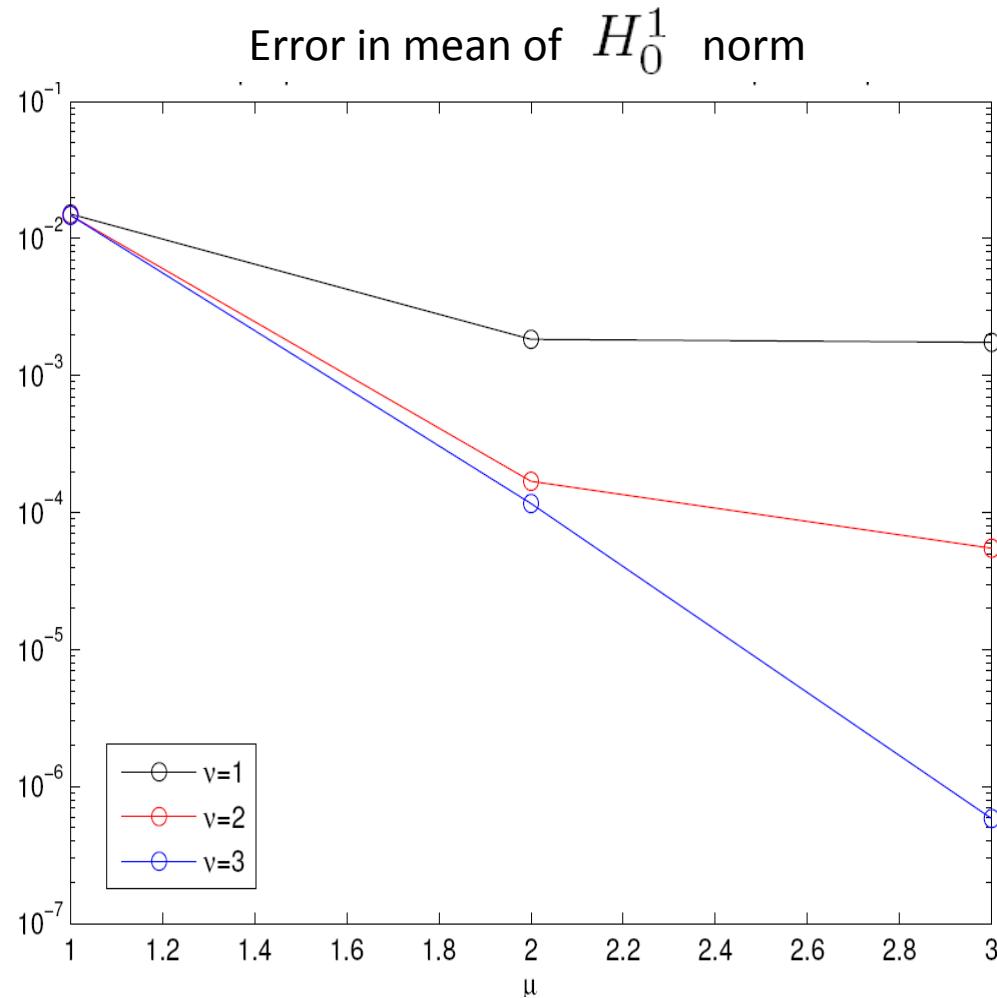
$\nu$  : dimension of each sub-problem

$\mu$  : order of approximation in each dimension

Use of ANOVA-type decompositions for standard PCM - Griebel, Schwab (2005, 2007).

# MEPCM-A convergence studies

2D Stochastic elliptic problem with 50D random input

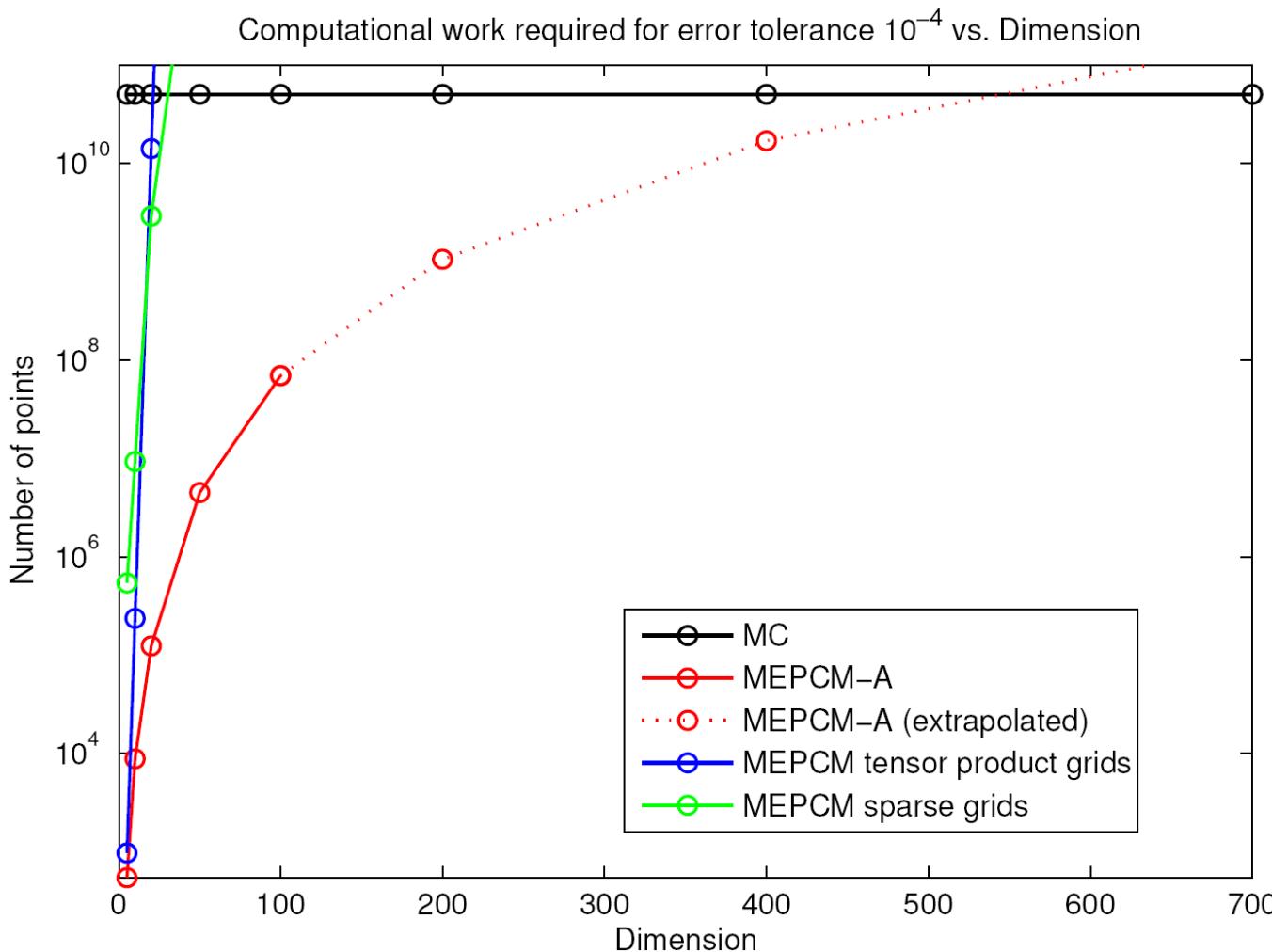


- Random input: random process w/Gaussian covariance kernel
- 50 - dimensional K-L input, 2 spatial dimensions
- MEPCM-A first two dimensions discretized
- Gauss-Legendre grids in each element

# MEPCM-A: high dimensional integration

The GENZ function  $f_6$  (DISCONTINUOUS) is defined as:

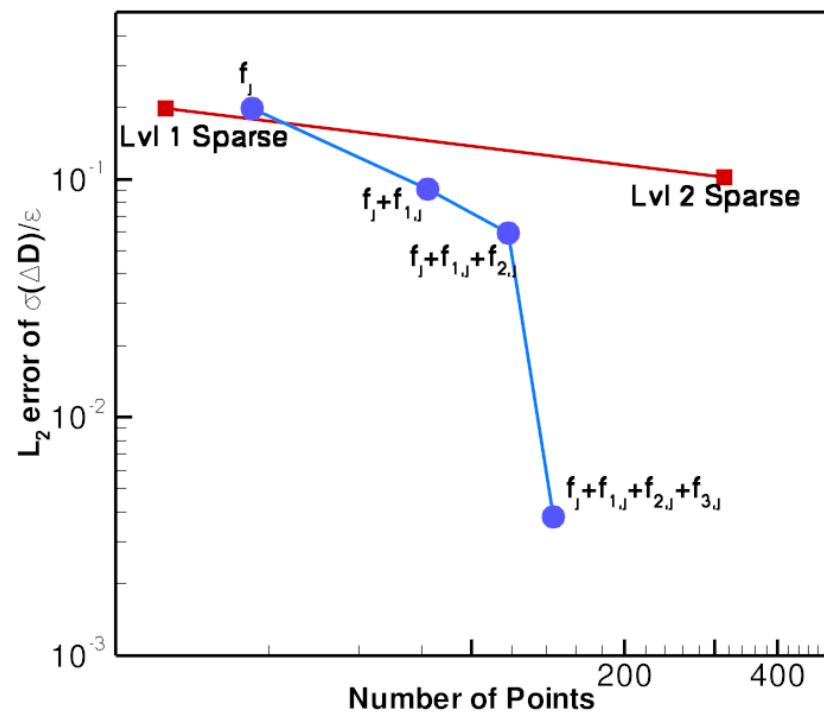
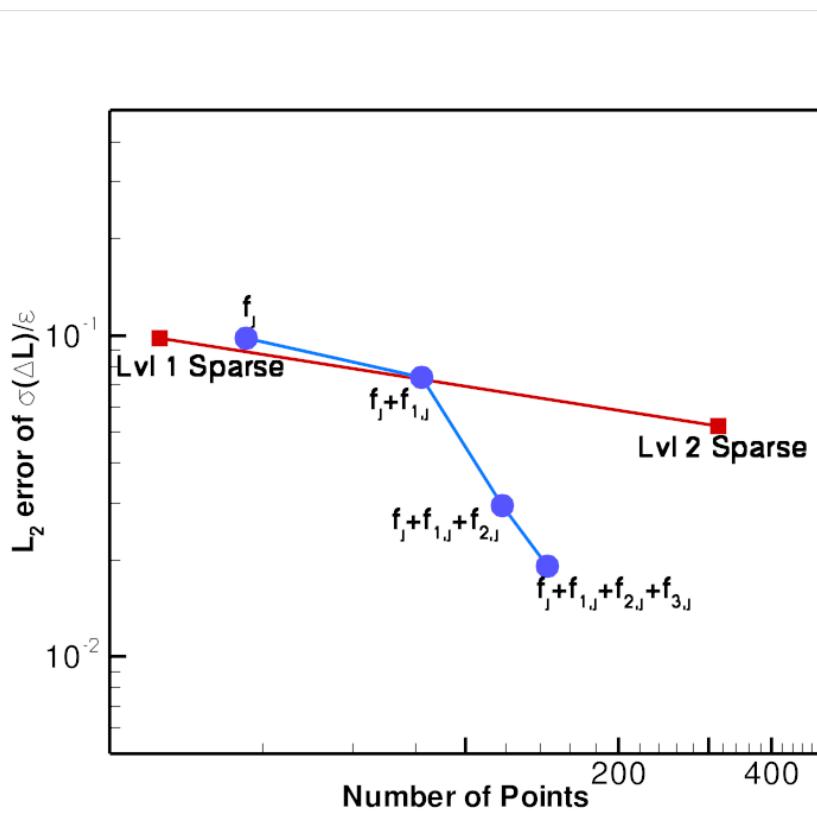
$$f_6(\mathbf{x}) = \begin{cases} 0, & \text{if } x_1 > w_1 \text{ or } x_2 > w_2 \\ \exp \sum_{i=1}^N c_i x_i, & \text{otherwise} \end{cases}$$



- Coefficients chosen to decay exponentially
- MEPCM/MEPCM-A meshes adapted to discontinuities
- MEPCM-A, tensor Gauss-Legendre grids
- MEPCM sparse - Smolyak Clenshaw Curtis
- PCM/PCM-A methods do not converge (sparse or tensor product)

# $L_2$ Error of Sparse Grid and Adaptive ANOVA Method

Mach number 8, correlation length 0.1, roughness length 1, nominal dimension 12, effective dimension 6,  $\sigma = 3$ . Error of standard deviation of extra lift (left) and drag (right).



# Stochastic Sensitivity Analysis

## Motivation:

- Rank all inputs and parameters in order of their significance to output variation
- Reduce dimension of parametric space in experiments or simulations

## Sensitivity Algorithms:

- Approximated Gradient Method

Morris, QMC, MC, Multi-Element Sparse Collocation

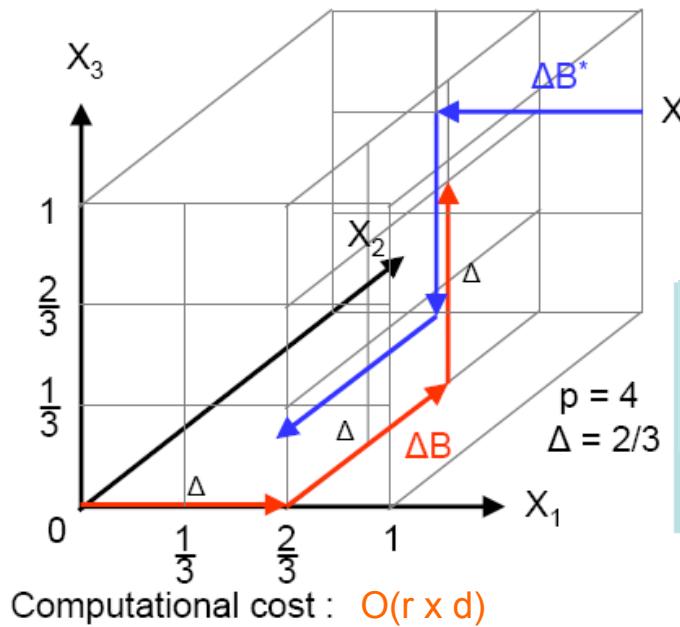
$$EE_i^j(x_1^0, \dots, x_d^0) = \frac{|y_j(x_1^0, x_2^0, \dots, x_{i-1}^0, x_i^0 + \Delta, x_{i+1}^0, \dots, x_d^0) - y_j(x_1^0, \dots, x_d^0)|}{\Delta}$$

Input  $X_i : i=1:d$   
Output  $y_j : j=1:n$



Pacific Northwest  
NATIONAL LABORATORY  
21

Lin & Karniadakis, AIAA-2008-1073, 2008; IJNME 2009



## Morris method (Morris, 1991)

**Objective :** efficiently identify sensitive parameters in system with many parameters

Divide range of uncertainty factor into equally spaced grid of  $p$  level, then compute incremental ratios:

$$X_i \rightarrow X_i + \Delta ; \Delta = p / 2 (p - 1)$$

Compute Elementary Effect (EE) from randomized trajectories:

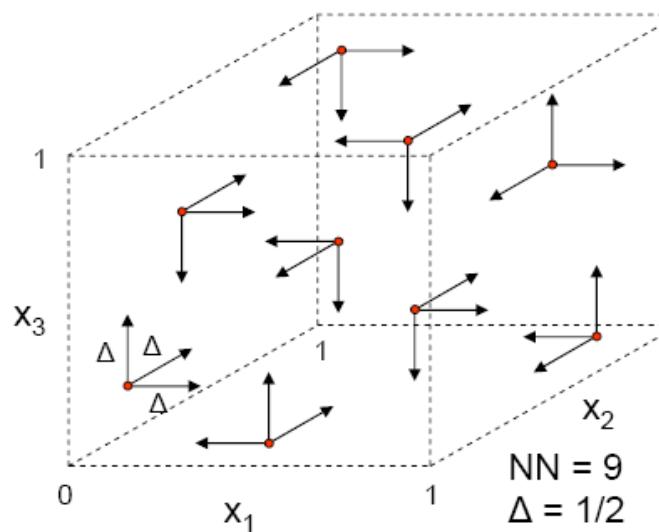
$$B^* = \left( J_{d+1,1} x^* + \left( \frac{\Delta}{2} \right) [(2B - J_{d+1,d}) D^* + J_{d+1,d}] \right) P^* ; B^* = \begin{pmatrix} 1 & 1 & 2/3 \\ 1/3 & 1 & 2/3 \\ 1/3 & 1 & 0 \\ 1/3 & 1/3 & 0 \end{pmatrix} ; B = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Perform  $NN$  realizations of  $EE_i [EE_{(i,1)}, EE_{(i,2)}, \dots, EE_{(i,NN)}]$  at  $NN$  different initial points  $X_1, \dots, X_{NN}$ , and then calculate mean (first-order effects) and standard deviation (coupling and nonlinear effects)

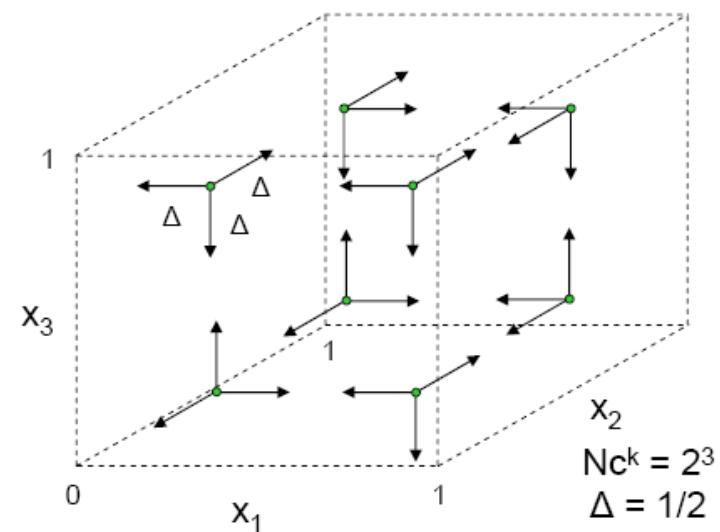
## Monte Carlo Sampling Method

## Collocation Method

**Objective :** efficiently identify sensitive parameters in system with many parameters without generating randomized trajectories



Cost of MC Sampling :  $O( NN \times d )$



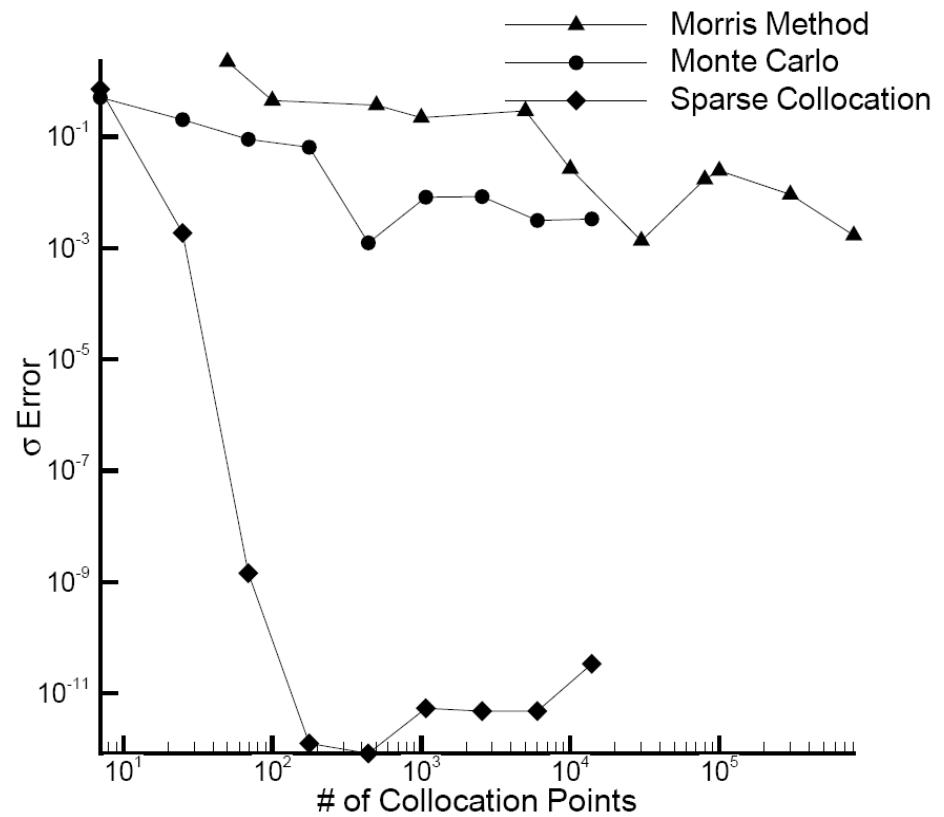
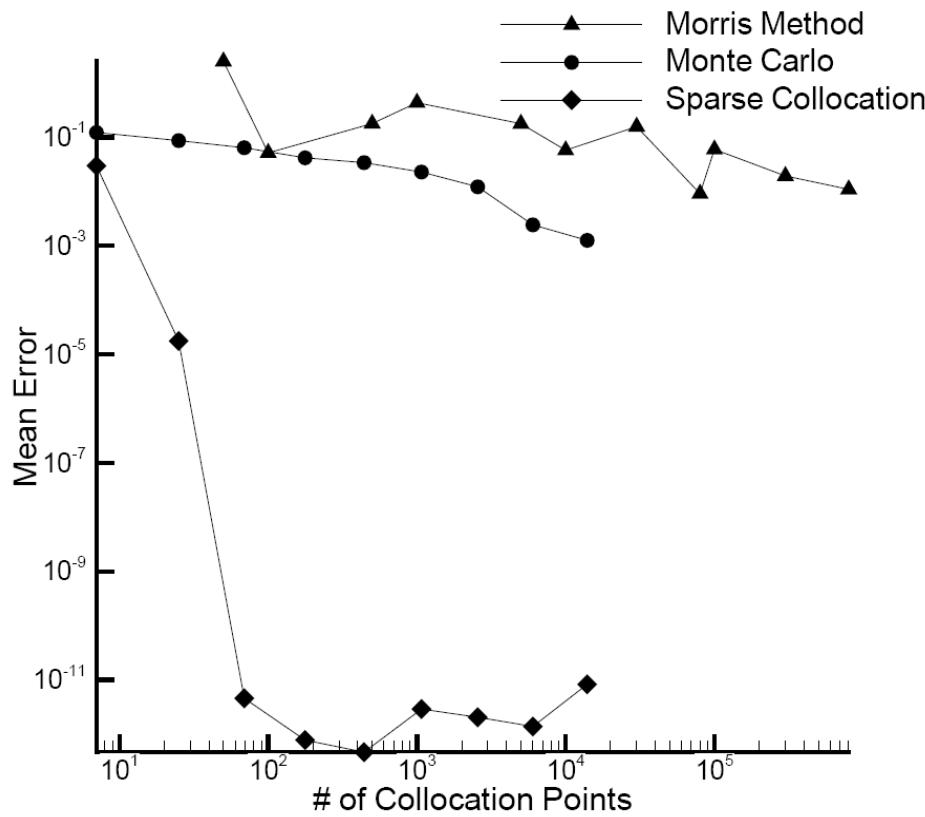
Cost of Collocation :  $O( Nc^d \times d )$

Calculate mean (first-order effects) and standard deviation (coupling and nonlinear effects) of  $EE_i$  [ $EE_{(i,1)}, EE_{(i,2)}, \dots, EE_{(i,NN)}$ ] at  $NN$  initial points for each  $i$  input

# Convergence Study of Stochastic Sensitivity Methods

$$y = 63e^{4x_1} - 70e^{3x_2} + 15e^{2x_3}$$

$$\text{Mean error} = \frac{|E_{\text{num}}[d_i^j] - E_{\text{ext}}[d_i^j]|}{|E_{\text{ext}}[d_i^j]|}, \quad \sigma \text{ error} = \frac{|\sigma_{\text{num}}(d_i^j) - \sigma_{\text{ext}}(d_i^j)|}{|\sigma_{\text{ext}}(d_i^j)|}$$



Lin & Karniadakis, AIAA-2008-1073, 2008; IJNME 2009

# Open Issues

## 1 Parametric Discontinuities/Bifurcations

- ❖ Multi-Element generalized Polynomial Chaos (ME-gPC)
- ❖ Multi-Element PCM (ME-PCM)

## 2 High-Dimensions

- ❖ Sparse Grids
- ❖ ME-PCM with ANOVA
- ❖ Sensitivity Analysis based UQ method

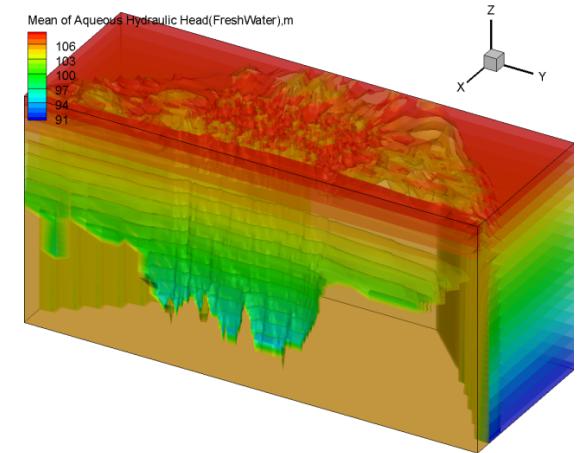
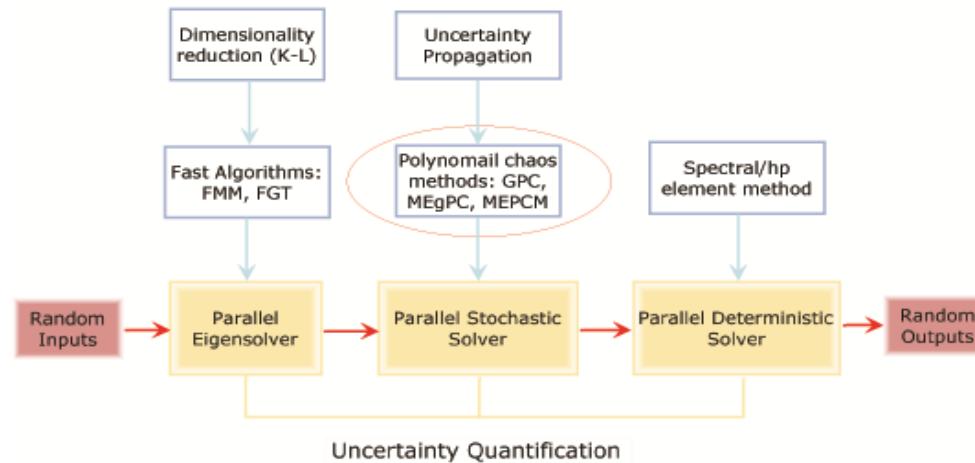
## 3 Scalability on Exascale Platforms

- ❖ Multilevel Parallelism for UQ at the Exascale
- ❖ Stochastic Mutigrid Method

## 4 Uncertainty Quantification in Random Composites

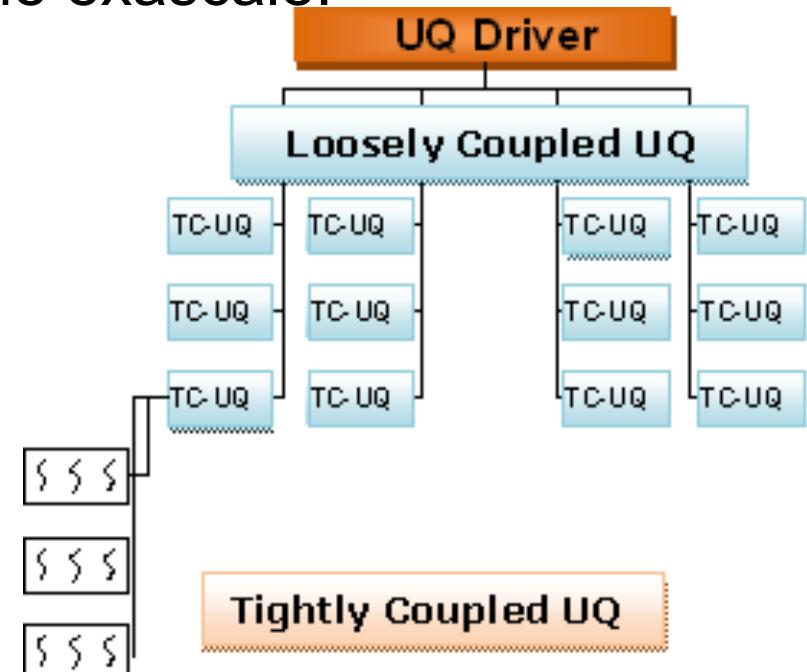
- ❖ Combine random domain decomposition with multi-element probabilistic collocation method

## Open Issue 3: Exascale Challenge - Scalable UQ Algorithm Design at the Exascale



### ► Multilevel Parallelism for UQ at the exascale:

- Upper level – Loosely coupled UQ
- Lower level – Tightly coupled UQ



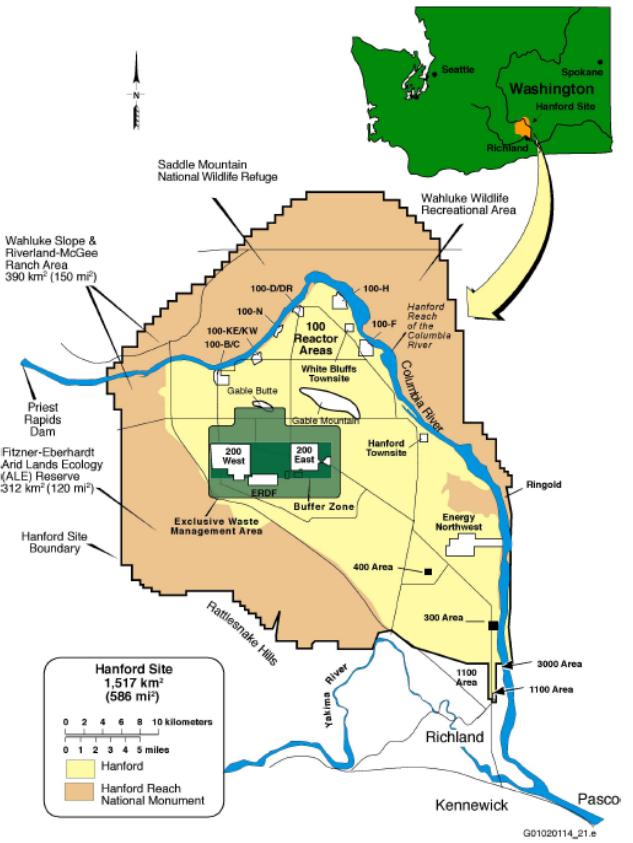
Pacific Northwest  
NATIONAL LABORATORY

# Uncertainty Quantification with Minimum Computational Cost

Dimension Reduction of Random Fields:  
**Karhunen-Loeve Decomposition**

Speed up Stochastic Simulations for Fast Convergence :  
**Probabilistic Collocation Methods on Sparse Grids**

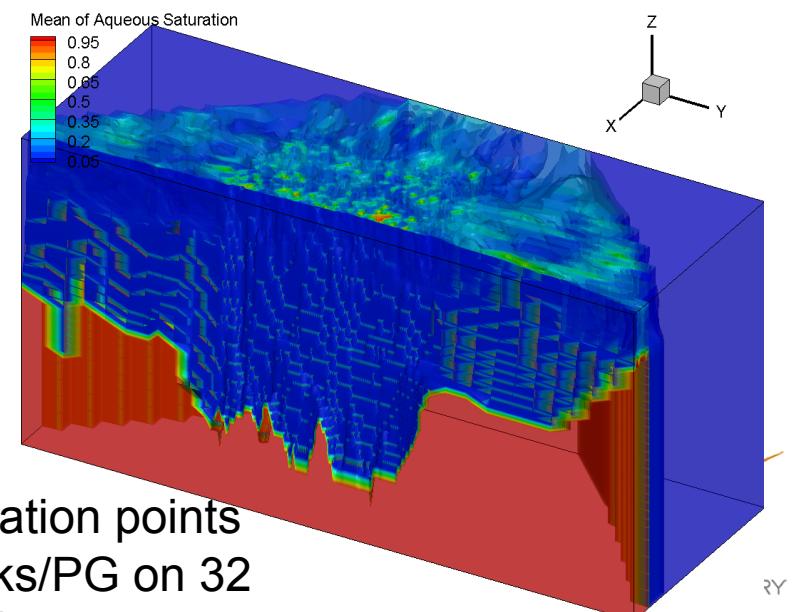
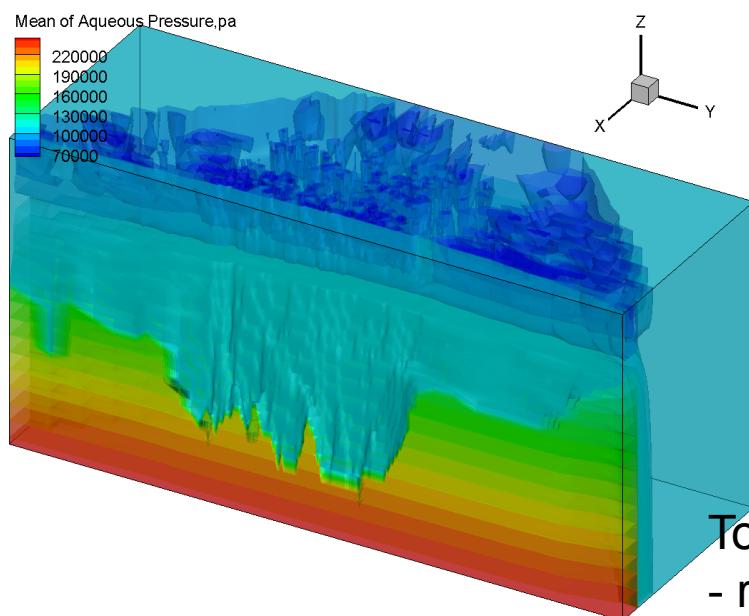
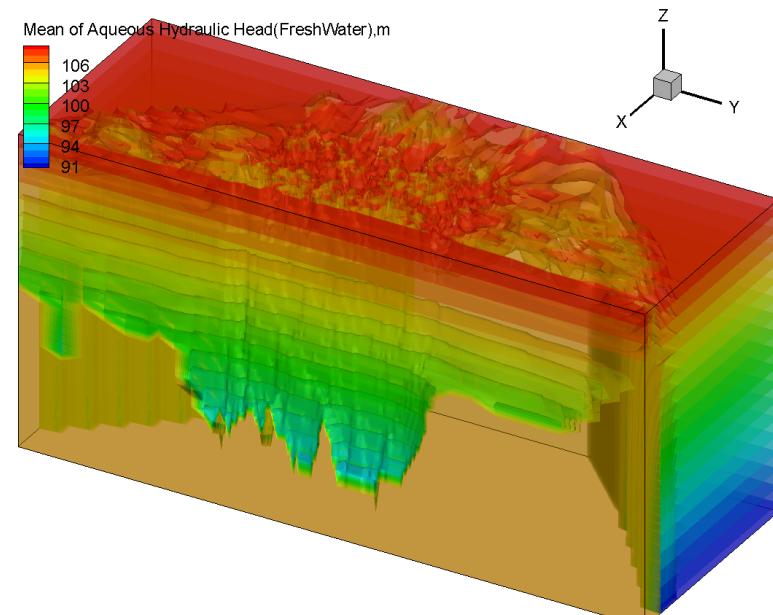
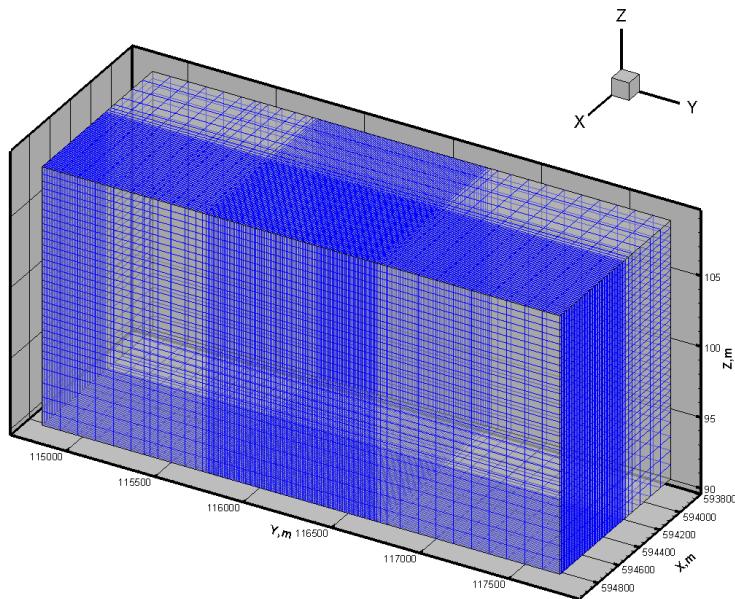
Reduce Communicate Cost: **Multilevel Parallelism using Processor Groups**



Provide Error Bar for Subsurface Flow and Reactive Transport

Confidence Bounds and Risk Assessment for Decision Makers

# Uncertainty Quantification of 3D Subsurface Flow in the Hanford 300 Area (Hydraulic Conductivity is a random field)



Total 317 Collocation points  
- running 10 tasks/PG on 32  
PG on 4096 nodes

# Open Issues

## 1 Parametric Discontinuities/Bifurcations

- ❖ Multi-Element generalized Polynomial Chaos (ME-gPC)
- ❖ Multi-Element PCM (ME-PCM)

## 2 High-Dimensions

- ❖ Sparse Grids
- ❖ ME-PCM with ANOVA
- ❖ Sensitivity Analysis based UQ method

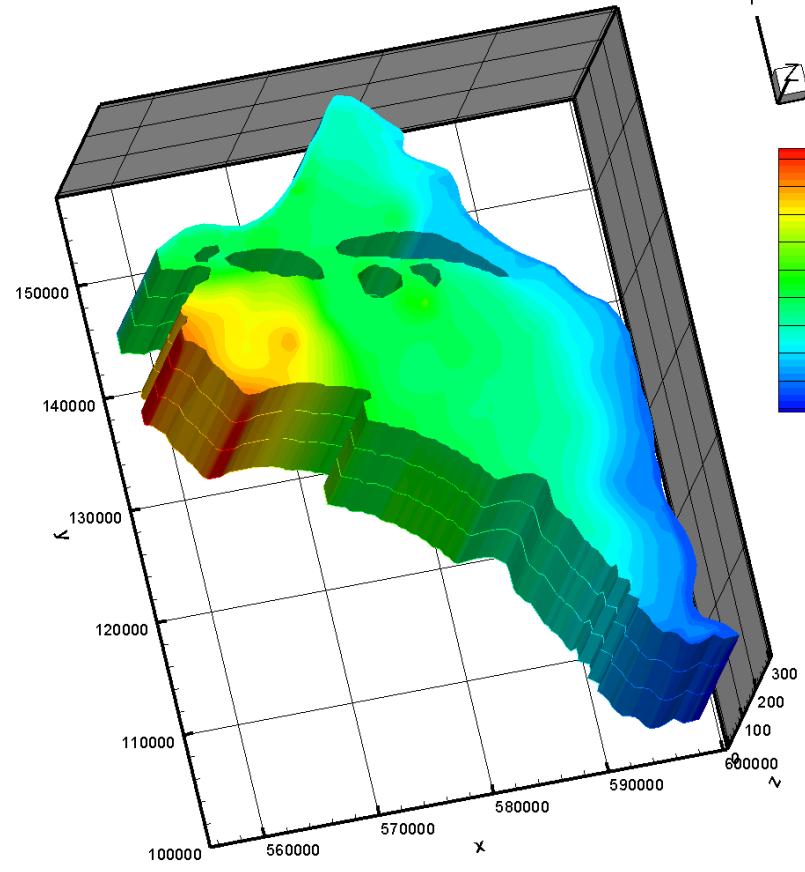
## 3 Scalability on Exascale Platforms

- ❖ Multilevel Parallelism for UQ at the Exascale
- ❖ Stochastic Mutigrid Method

## 4 Uncertainty Quantification in Random Composites

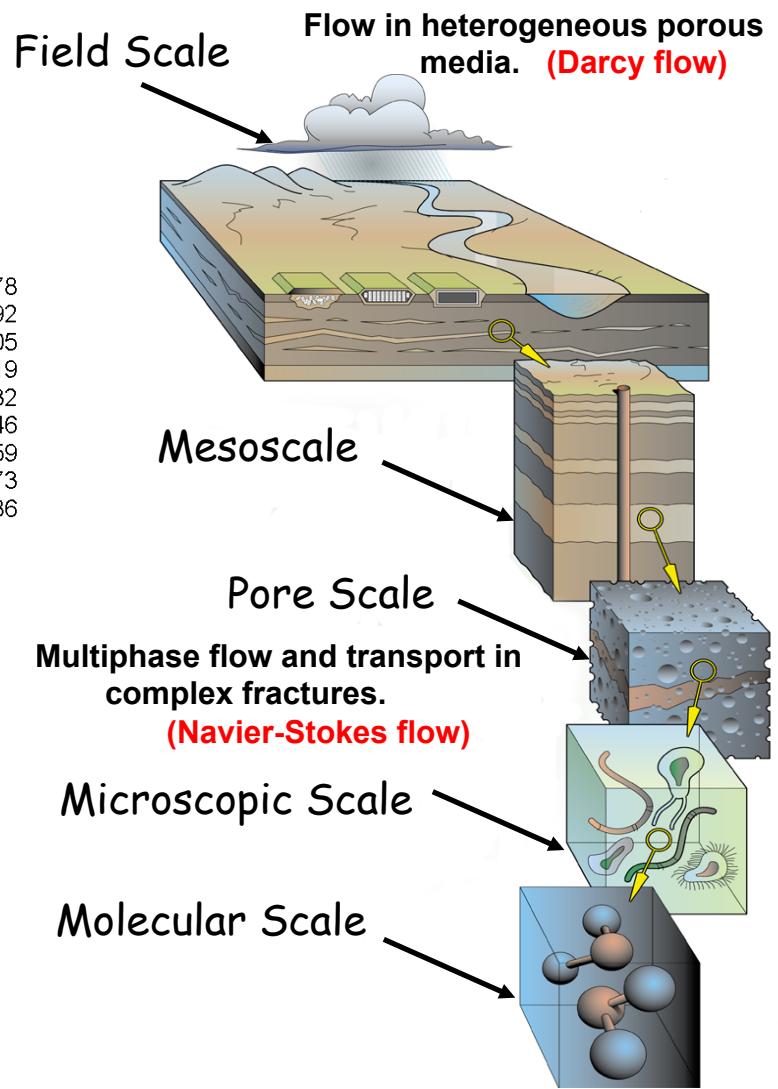
- ❖ Combine random domain decomposition with multi-element probabilistic collocation method

# Motivation:



Hydraulic Head Mean

Multilayer heterogeneous  
porous media



# Stochastic collocation method on sparse grids for flow and transport in randomly heterogeneous porous media

## Governing Equations

$$\nabla^* \cdot [\kappa_s^*(\mathbf{x}^*; \omega) \nabla^* h^*(\mathbf{x}; \omega)] = 0,$$

$$\frac{\partial C^*}{\partial t^*} + \nabla^* \cdot (v^* \cdot C^*) = \nabla^* \cdot (D_f^* \nabla^* C^*) = \nabla^* \cdot [(D_w^* + \alpha^* |v^*|) \nabla^* C^*],$$

subject to initial and boundary conditions:

$$h^*(\mathbf{x}^*) = H(\mathbf{x}^*), \quad \mathbf{x}^* \in \Gamma_D^*,$$

$$\frac{\partial C^*(\mathbf{x}^*, t^*; \omega)}{\partial X^*} = 0, \quad \mathbf{x}^* \in D^*,$$

$$\frac{\partial C^*(\mathbf{x}^*, t^*; \omega)}{\partial X^*} = 0, \quad \mathbf{x}^* \in \Gamma_D^*,$$

## Open Issue 4: highly heterogeneous random composites - Combining Random Domain Decomposition & Probabilistic Collocation Method for Random Composites

Random composites with two dominant scales of uncertainty:

- ▶ Large-scale uncertainty in the spatial arrangement of materials
- ▶ Small-scale uncertainty in the parameters within each material

### Joint Probability Density Function

$$p_{\kappa}(k, \gamma) = p_{\kappa}(k|\gamma)p_{\tau}(\gamma),$$

$$\begin{aligned} < h(\mathbf{x}) > &= \int \int h(\mathbf{x}; k, \gamma) dk d\gamma = \int \int h(\mathbf{x}; k, \gamma) p(k|\gamma) p_{\tau}(\gamma) dk d\gamma, \\ &= \int < h(\mathbf{x}) | \gamma > p_{\tau}(\gamma) d\gamma, \\ &= \sum_{q=1}^Q < h(\mathbf{x}) | \alpha_q > \omega_q. \end{aligned}$$

Along the random boundary  $\alpha$ ,

$$h|_{x=\alpha^-} = h|_{x=\alpha^+}, \quad \kappa(\mathbf{x}) \frac{\partial h}{\partial x}|_{x=\alpha^-} = \kappa(\mathbf{x}) \frac{\partial h}{\partial x}|_{x=\alpha^+}.$$

[Lin et al., JCP, 2010](#)

## Analytical Solution for One-Dimensional Random Contact Point Problem

Consider a one-dimensional differential equation,

$$\frac{d}{dx} \left[ K(x) \frac{dh(x)}{dx} \right] = 0, \quad x \in (0, 1)$$

subject to the boundary conditions

$$K \frac{dh}{dx}(x = 0) = -q, \quad h(x = 1) = 0$$

The random coefficient  $K(x)$  is given by

$$K(x) = \begin{cases} K_1(x), & 0 < x < \alpha, \\ K_2(x), & \alpha < x < 1 \end{cases}$$

Mean,  $\langle h \rangle = \int_D \langle h | \alpha \rangle p_\tau(\alpha) d\alpha$ ,

$$\text{where } \frac{\langle h | \alpha \rangle}{q} = \mathcal{H}(\alpha - x) \left[ \frac{\alpha - x}{K_{h_1}} + \frac{1 - \alpha}{K_{h_2}} \right] + \mathcal{H}(x - \alpha) \frac{1 - x}{K_{h_2}}$$

Variance,  $\sigma_h^2(x) = \langle h^2 \rangle - \langle h \rangle^2$ , where

$\langle h^2 \rangle = \int_D \langle h^2 | \alpha \rangle p_\tau(\alpha) d\alpha$  and

$$\begin{aligned} \frac{\langle h^2 | \alpha \rangle}{q^2} &= \mathcal{H}(\alpha - x) \left[ \frac{1}{K_{h_1}^2} \int_x^\alpha \int_x^\alpha e^{\sigma_{Y_1}^2 \rho_{Y_1} (s_1 - s_2)} ds_1 ds_2 + \frac{2(\alpha - x)(1 - \alpha)}{K_{h_1} K_{h_2}} \right. \\ &\quad \left. + \frac{1}{K_{h_2}^2} \int_\alpha^1 \int_\alpha^1 e^{\sigma_{Y_2}^2 \rho_{Y_2} (s_1 - s_2)} ds_1 ds_2 \right] + \mathcal{H}(x - \alpha) \frac{1}{K_{h_2}^2} \int_x^1 \int_x^1 e^{\sigma_{Y_2}^2 \rho_{Y_2} (s_1 - s_2)} ds_1 ds_2 \end{aligned}$$

## Boundary and initial conditions for 1D Random contact point & inclusion:

$$f = -k \frac{\partial h}{\partial x} = 1, \frac{\partial C}{\partial x} = 0 \quad \xrightarrow{\alpha = U(2.7, 3.3)} \quad h = 0, \frac{\partial C}{\partial x} = 0$$

**Along the random boundary  $\alpha$ ,**

$$h|_{x=\alpha^-} = h|_{x=\alpha^+}, \quad \kappa(\mathbf{x}) \frac{\partial h}{\partial x}|_{x=\alpha^-} = \kappa(\mathbf{x}) \frac{\partial h}{\partial x}|_{x=\alpha^+}.$$

$$f = -k \frac{\partial h}{\partial x} = 1, \frac{\partial C}{\partial x} = 0 \quad \xrightarrow{\quad | \quad | \quad | \quad} \quad h = 0, \frac{\partial C}{\partial x} = 0$$

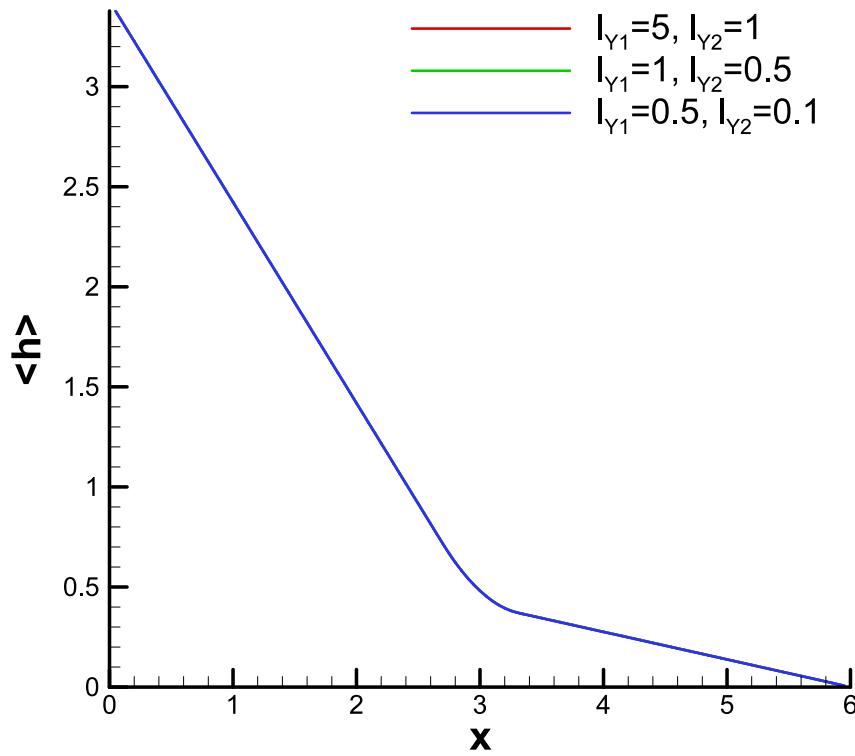
**Along the random boundary  $\alpha \pm 1$ ,**

$$h|_{x=(\alpha\pm 1)^-} = h|_{x=(\alpha\pm 1)^+}, \quad \kappa(\mathbf{x}) \frac{\partial h}{\partial x}|_{x=(\alpha\pm 1)^-} = \kappa(\mathbf{x}) \frac{\partial h}{\partial x}|_{x=(\alpha\pm 1)^+}.$$

*Initial Condition :*

$$C(x, 0) = \exp\left(\frac{(x - x_o)^2}{2\sigma_o^2}\right)$$

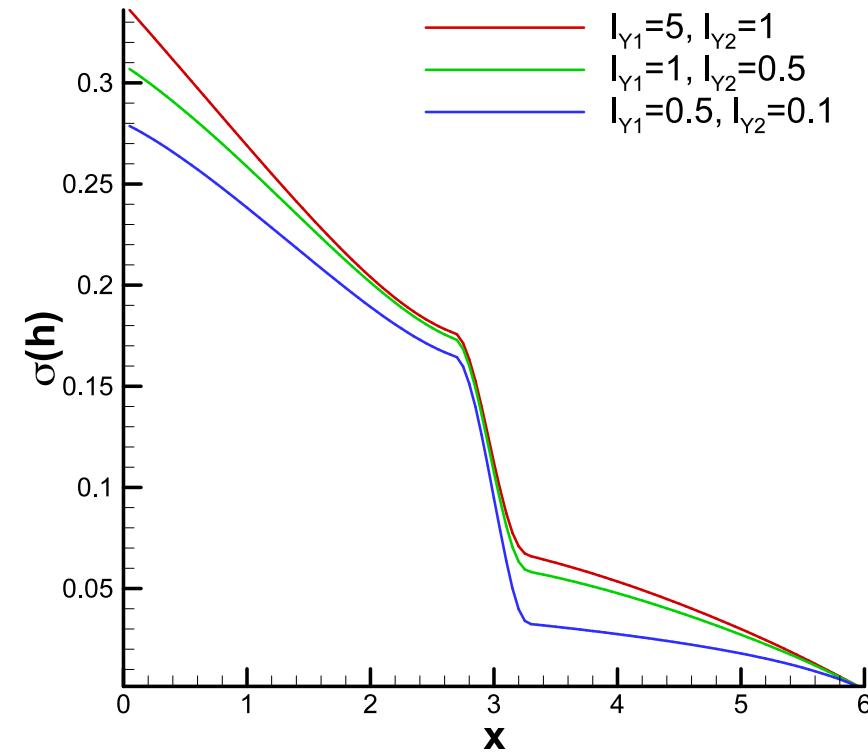
**1D Random contact point: The effect of correlation lengths of hydraulic conductivity on the mean and std of the hydraulic head for 1D random contact point problem.**



Mean of Head

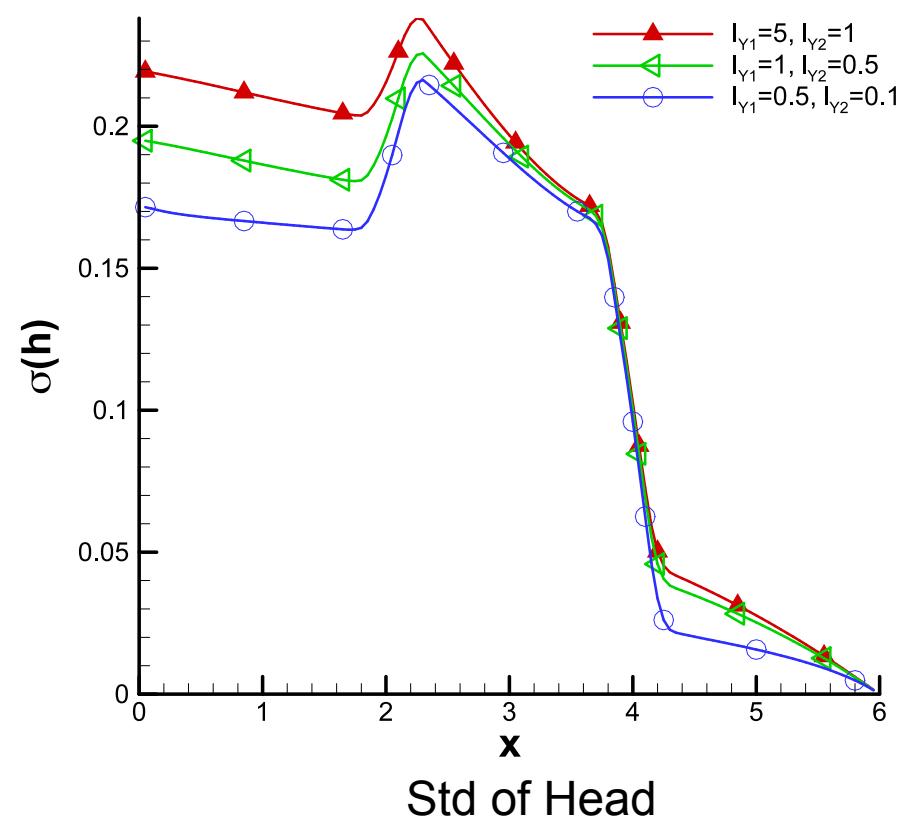
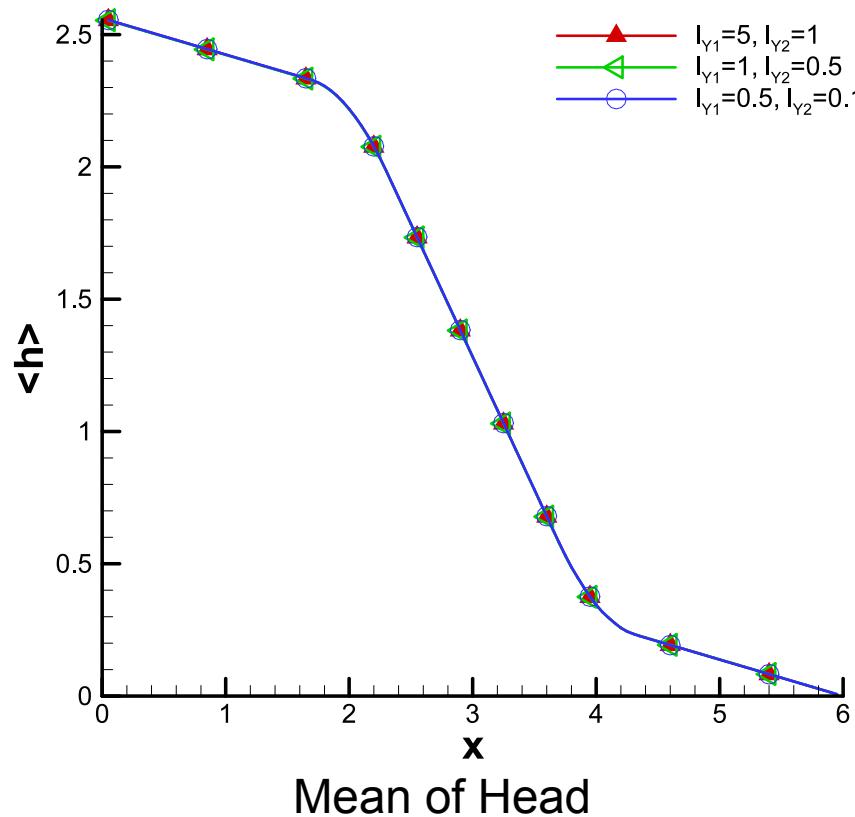
The boundary location  $\alpha$  between two different porous meidas is subject to Uniform distribution

[Lin et al., JCP, 2010](#)



Std of Head

# 1D Random Inclusion: The effect of correlation lengths of hydraulic conductivity on the mean and std of the hydraulic head for 1D random contact point problem.



The boundary location  $\alpha$  between two different porous media is subject to Uniform distribution [Lin et al., JCP, 2010](#)

# Uncertainty Quantification and Optimal Parameters Estimation in Convective Cloud scheme in Climate Model

► Ben Yang<sup>1,2</sup>, Yun Qian<sup>2</sup>, Guang Lin<sup>2</sup>, Ruby Leung<sup>2</sup>, and Yaocun Zhang<sup>1</sup>

► 1. Nanjing University, Nanjing, China

► 2. Pacific Northwest National Laboratory, WA, USA



## ► Five Parameters

Downdraft Rate  
Entrainment Rate  
CAPE Consumption Time  
TKE for Shallow Convection  
Starting Height of Downdraft

## ► Observational Constraint

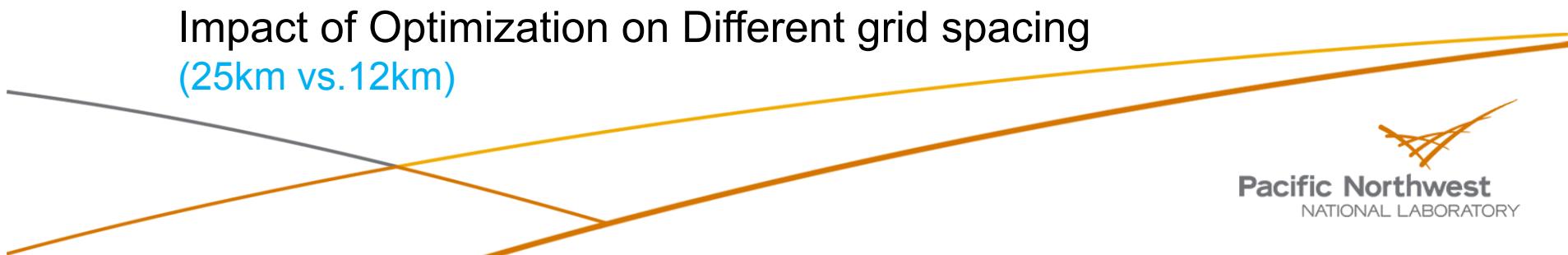
GPCP daily  
precipitation



(UW) Daily 1/8-degree gridded meteorological data  
(Maurer et al., 2002)

## ► Analysis

Impact of Optimization on Different grid spacing  
(25km vs. 12km)

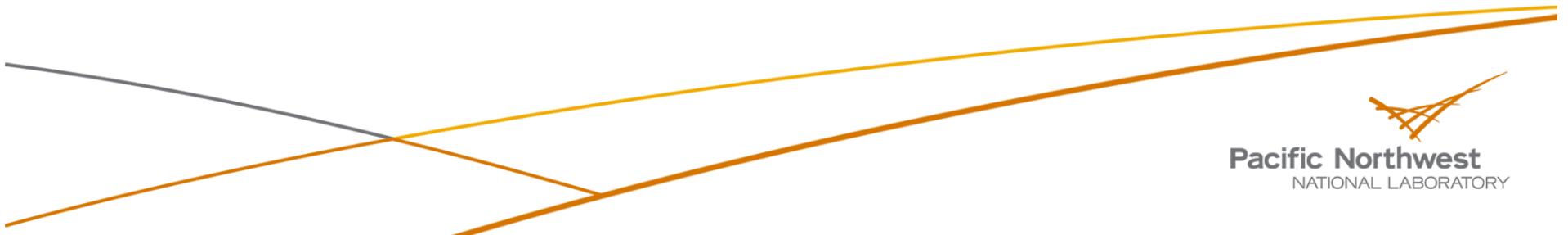


# Model uncertainty



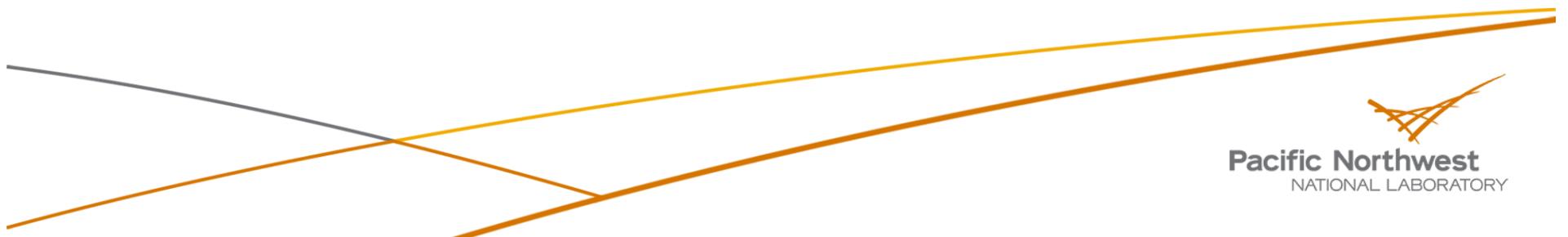
Unresolved physical processes  
Physical scheme  
Parameter values

- ▶ Models are always tuned towards the mean of observation
- ▶ A climate model's result depends nonlinearly to the combined changes in model parameters
- ▶ Parameter's errors may compensate with another parameter's errors



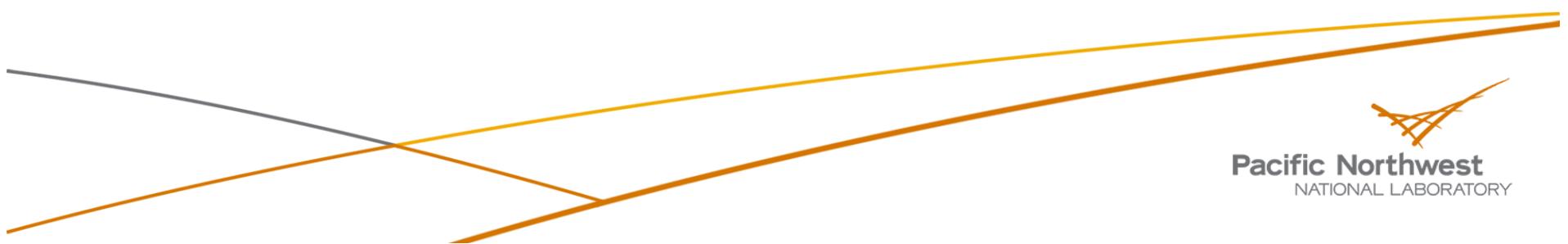
# Objective

- ▶ Quantifying the uncertainty and deriving **optimal parameters** used in cloud convection scheme in climate model
- ▶ Studying sensitivity of **other physic processes and simulations** to the parameters in cloud convection scheme



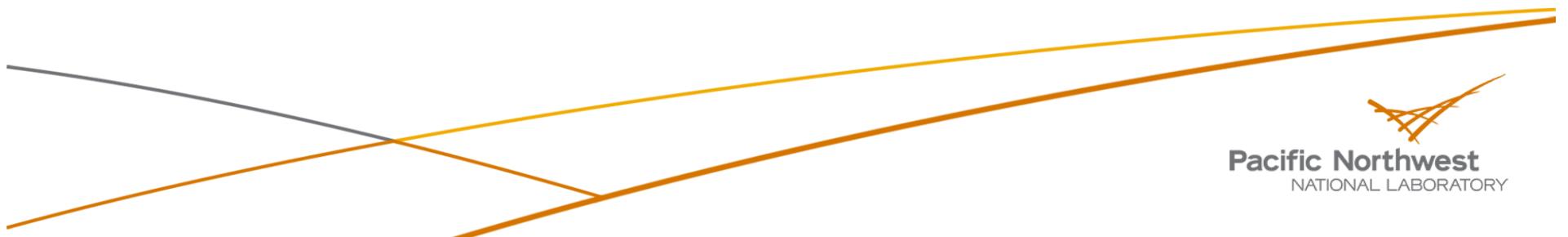
# Outline

- ▶ Optimize Method & Experiment Design
- ▶ Results
- ▶ Conclusion
- ▶ Next Step



# Optimized method & Experiment Design

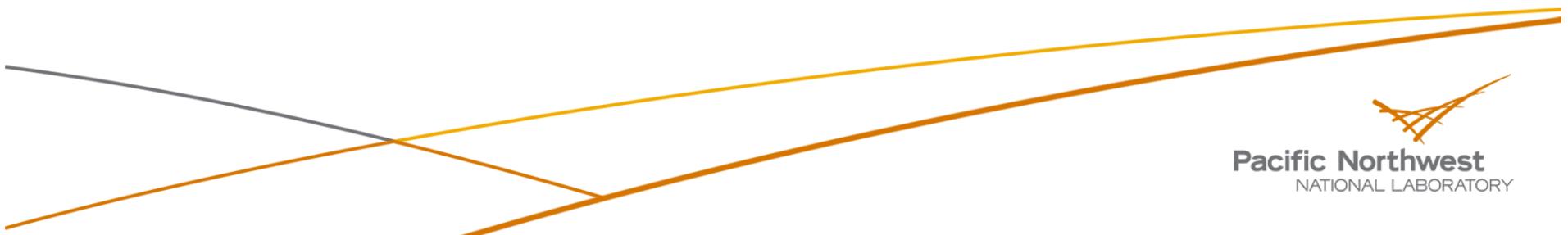
- ▶ Optimized Method
- ▶ Selected Parameters
- ▶ Model Configurations
- ▶ Observational Data



# Optimized Method

Identify a range of parameter sets that enable model predictions to be bounded within observational uncertainties

- ▶ Grid Search (**straightforward method**, Sen and Stoffa, 1996)
- ▶ Gibbs Sampler (**Metropolis algorithm**, Metropolis et al., 1953)
- ▶ Very Fast Simulated Annealing (**VFSA**, Ingber, 1989)
- ▶ Multiple VFSA (**MVFSA**, Jackson et al., 2008)



# VFSA

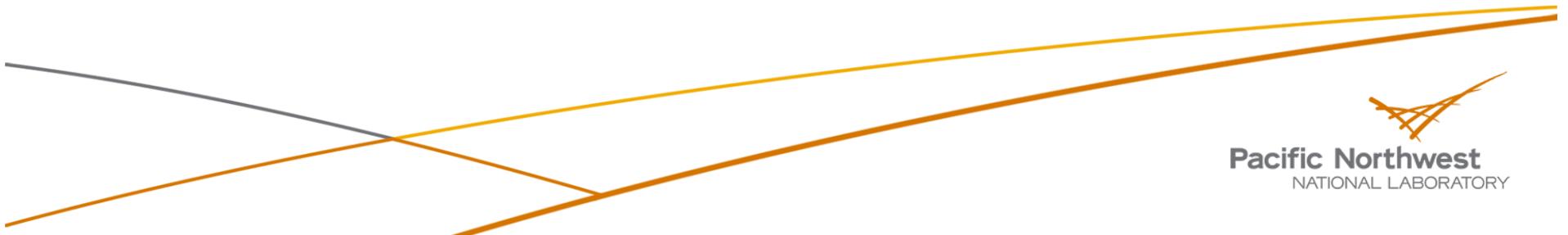
(Jackson et al, J. Climate, 2004)

$$m_i^{k+1} = m_i^k + y_i(m_i^{\max} - m_i^{\min})$$

$$y_i \in [-1,1]$$
$$m_i^{\max} \leq m_i^{k+1} \leq m_i^{\min}$$

$$y_i = \text{Sign}(Random - 0.5)T_k \left[ \left(1 + \frac{1}{T_k}\right)^{|2Random-1|} - 1 \right]$$

$$T_k = T_0 \exp[-\alpha(k-1)^{1/NM}]$$



# VFSA

(Jackson et al, J. Climate, 2004)

$E(m)$  :Evaluation Cost

$P(m)$  :Probability Density

N different sets of observation

$$E(m) = \sum_{i=1}^N \frac{1}{2N} \{ [d_{obs} - g(m)]^T \times C^{-1} [d_{obs} - g(m)] \}$$

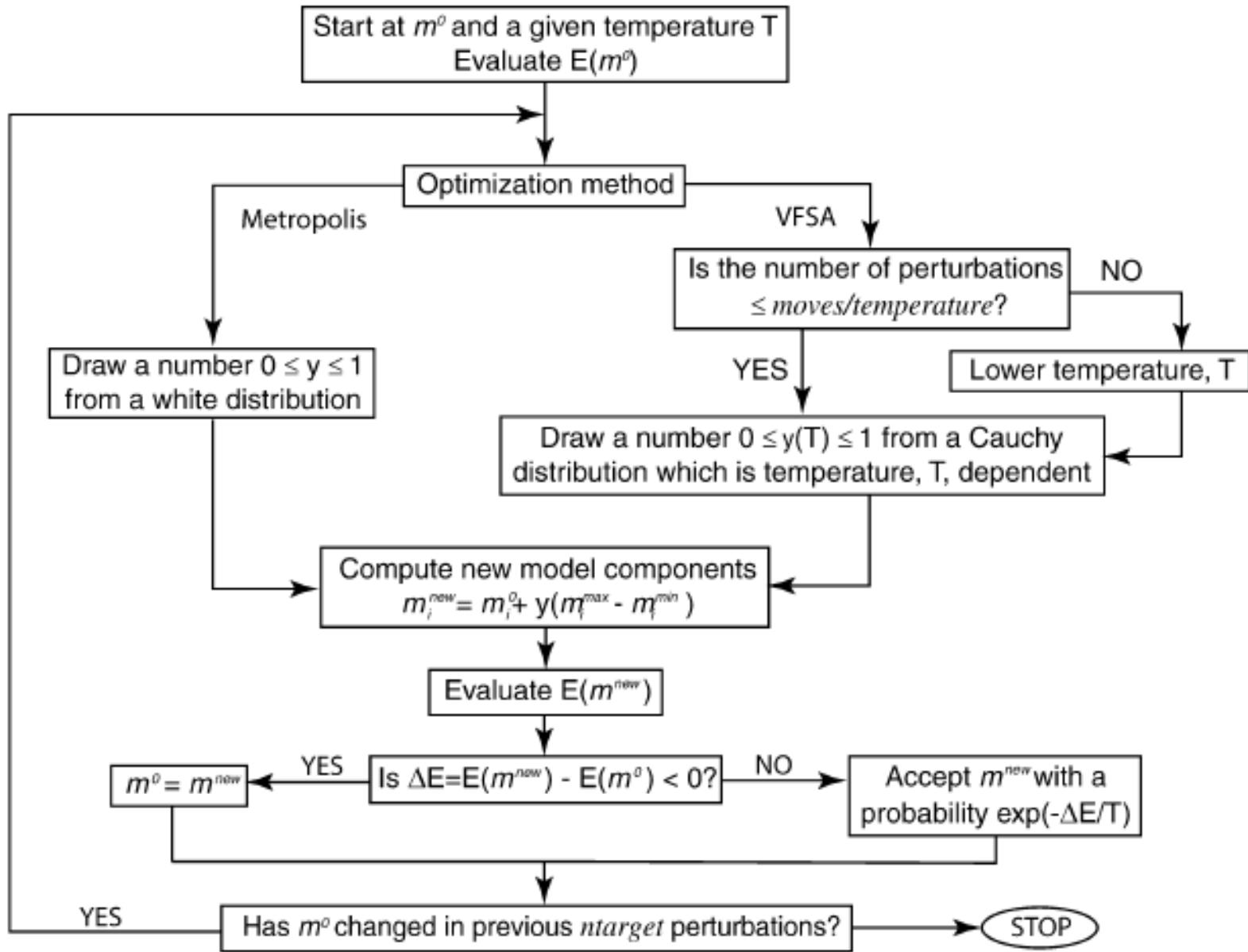
Observation      Simulation      Covariance Matrix

$$P(m) \propto \exp[-S \cdot E(m)]$$



# VFSA

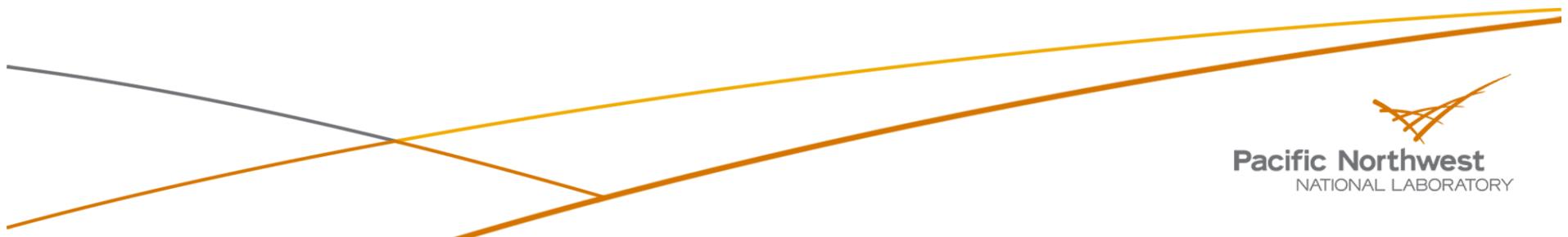
(Jackson et al, J. Climate, 2004)



est  
RATORY

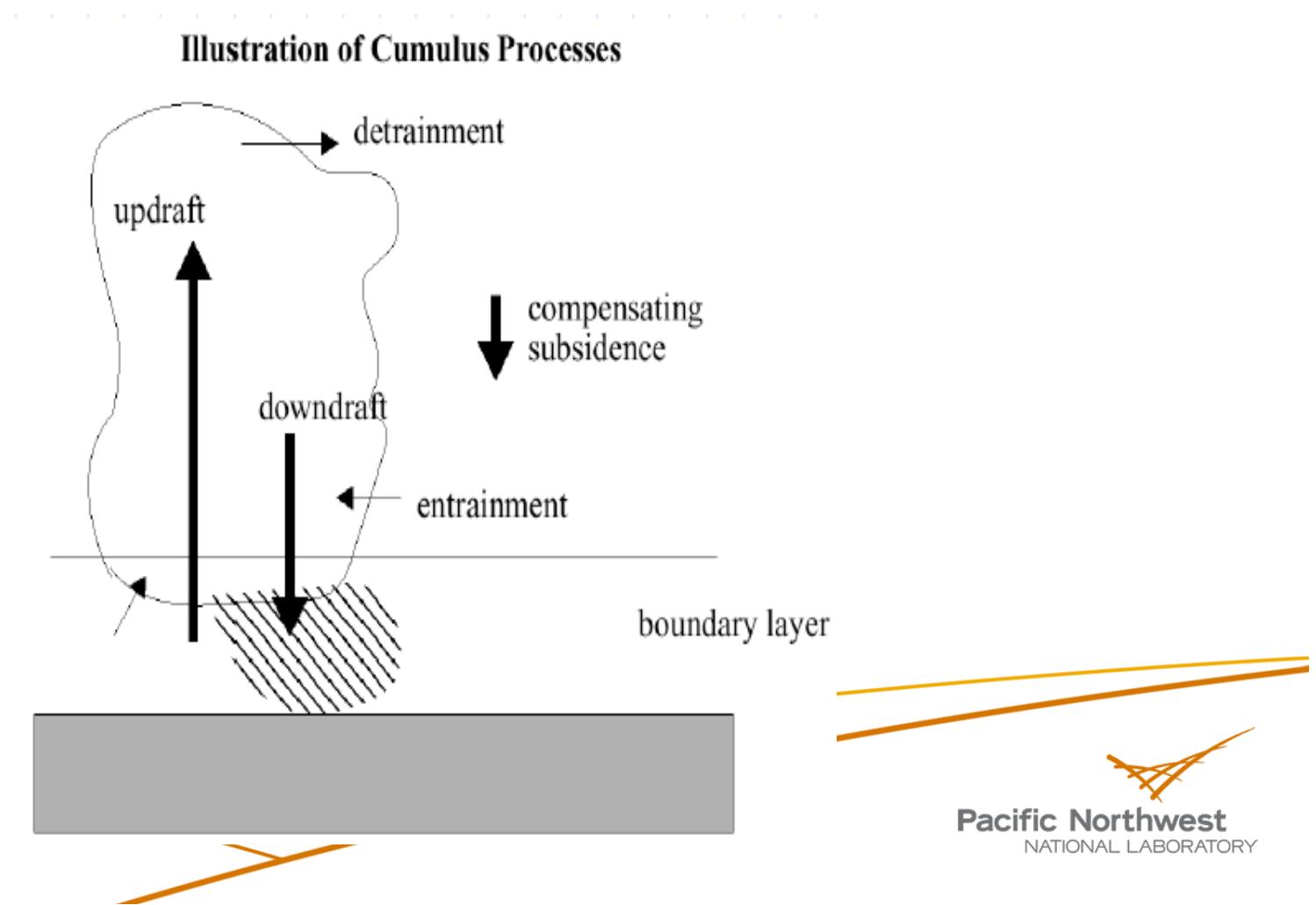
# Optimized method & Experiment Design

- ▶ Optimized Method
- ▶ Selected Parameters
- ▶ Model Configurations
- ▶ Observational Data



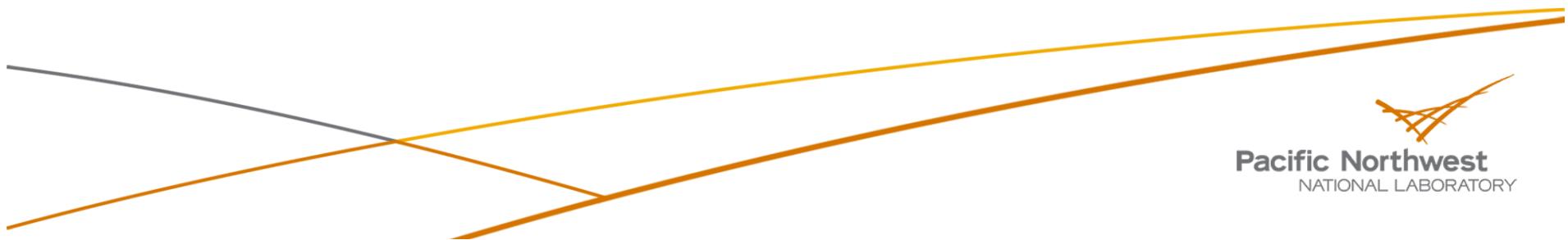
# Selected Parameters

Weather Research and Forecasting Model (**WRF3.2**)  
New Kain-Fritsch Cumulus scheme



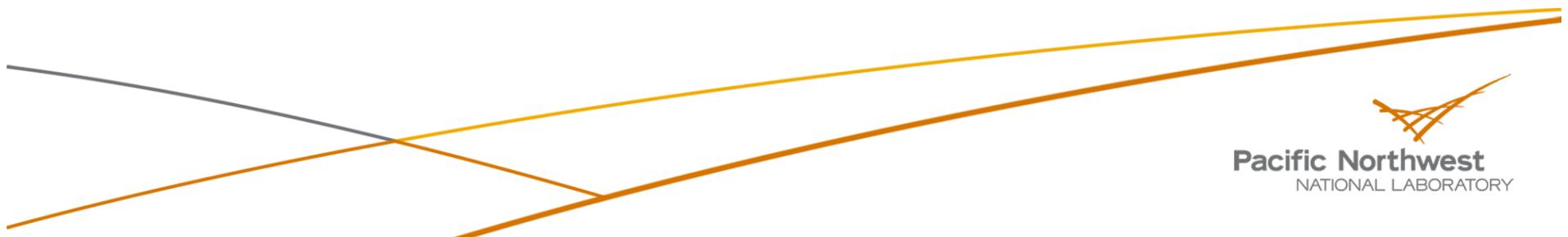
# Selected Parameters

| Parameter | Default | Minimum | Maximum | Description  |
|-----------|---------|---------|---------|--|
| Pd        | 0       | -1      | 1       | coefficient about downdraft mass flux rate                     |
| Pe        | 0       | -1      | 1       | coefficient about Entrainment mass flux rate                   |
| Pt        | 5       | 3       | 12      | maximum TKE in sub-cloud layer ( $\text{m}^2 \text{ s}^{-2}$ ) |
| Ph        | 150     | 50      | 350     | starting height of downdraft above cloud base (hPa)            |
| Pc        | 2700    | 900     | 7200    | average consumption time of CAPE (s)                           |



# Optimized method & Experiment Design

- ▶ Optimized Method
- ▶ Selected Parameters
- ▶ Model Configurations
- ▶ Observational Data



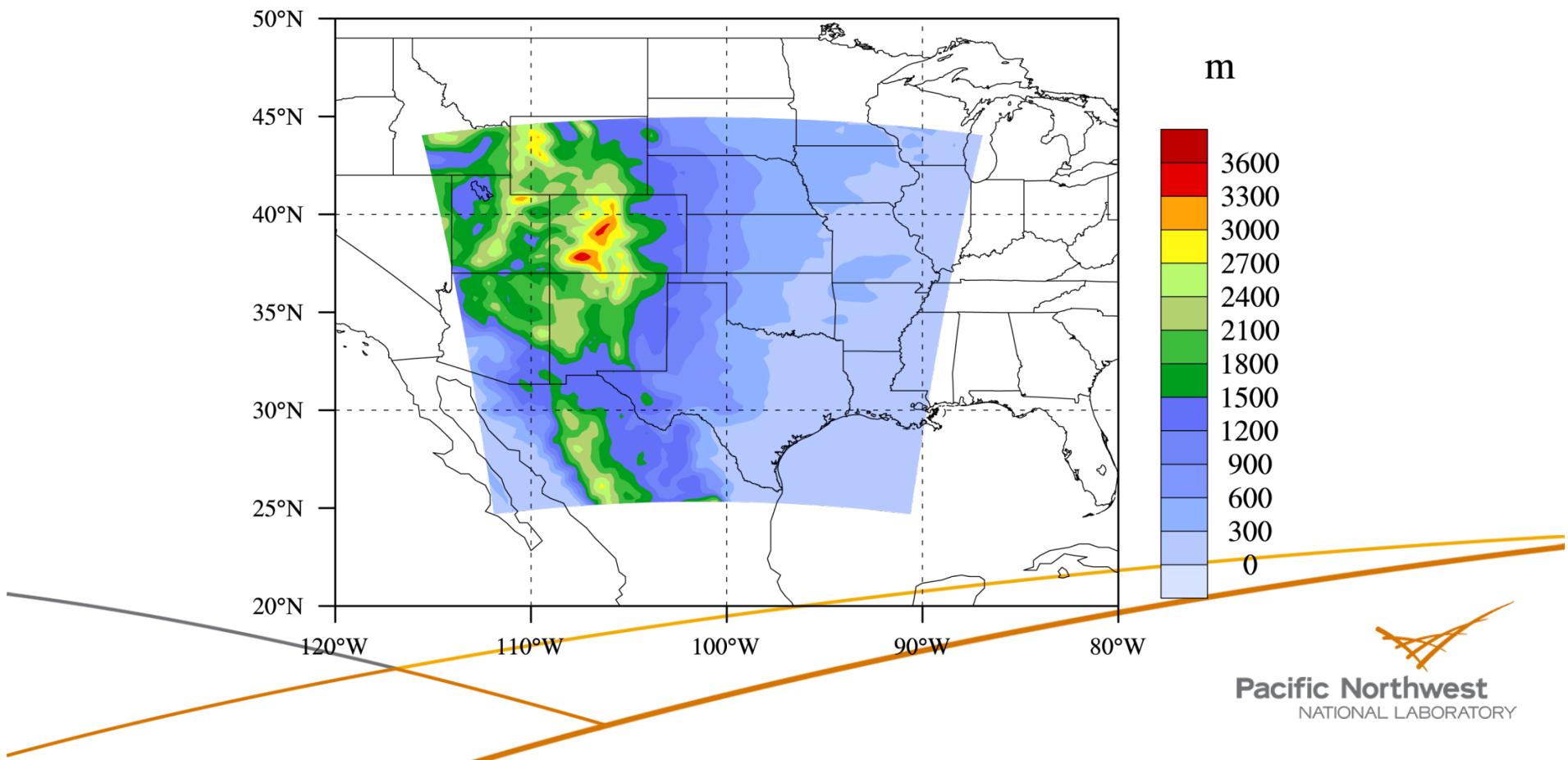
# Model Configuration

South Great Plain ( $25^{\circ}\text{N}$ - $44^{\circ}\text{N}$ ,  $112^{\circ}\text{W}$ - $90^{\circ}\text{W}$ )

Simulation period: May 1 to July 1, 2007

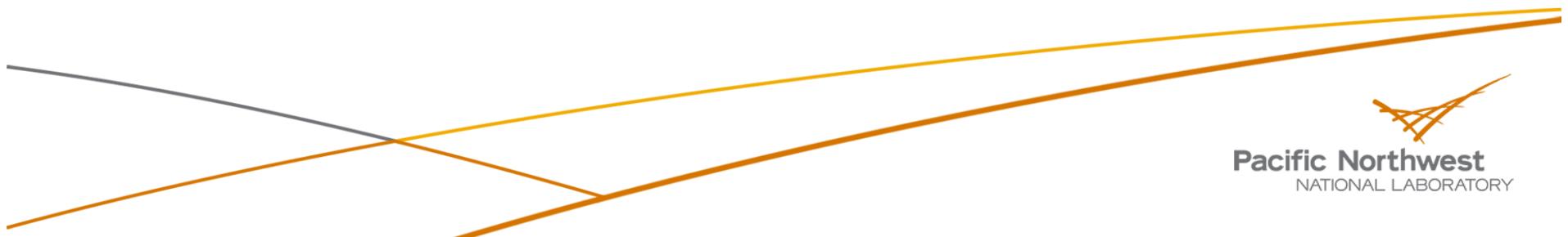
Analysis: June 2007

Terrain Height



# Optimized method & Experiment Design

- ▶ Optimized Method
- ▶ Selected Parameters
- ▶ Model Configurations
- ▶ Observational Data

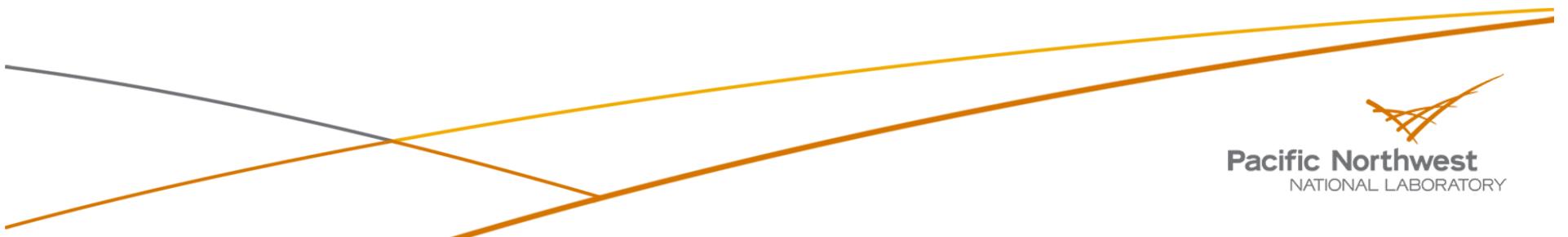


# Observational Data 1

- For Initial and Boundary Condition of WRF  
NARR Reanalysis data (32km-resolution, 3-hour interval)
  
- For Observational Constraint  
UW 1/8-degree gridded meteorological data: Daily Precipitation  
(Maurer et al., 2002)

$$E(m) = \left( \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K [d_{obs} - g(m)]^2 \right) / (I * J * K)$$

i, j: index of horizontal grid  
k: index of day



# Observational Data 2

- For Result Analysis

UW 1/8-degree gridded meteorological data:

Daily precipitation (mm/day), maximum and minimum temperature (C), and wind speed (m/s)

NARR Reanalysis data

$$E(m) = \left( \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K [d_{obs} - g(m)]^2 \right) / (I * J * K)$$

$$C(m) = \left( \sum_{k=1}^K SC[d_{obs}, g(m)] \right) / K$$



Spatial Correlation between simulation and observation

i, j: index of horizontal grid

k: index of day

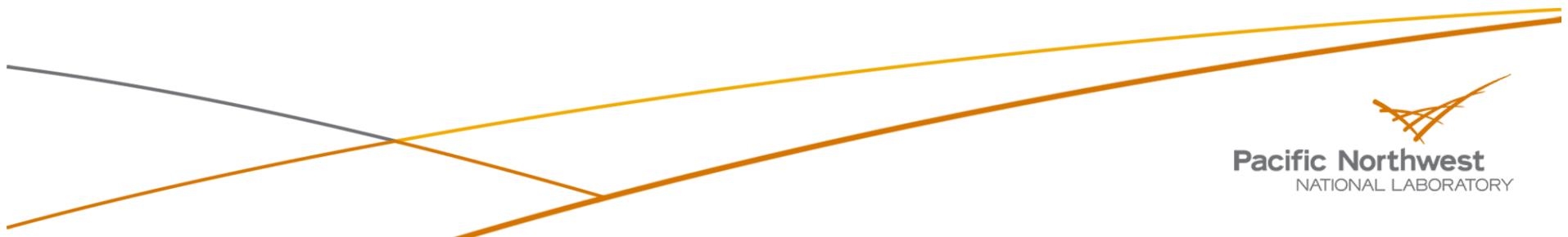
$$EM = E(m)$$

$$EC = -C(m)$$

Pacific Northwest  
NATIONAL LABORATORY

# Simulation Results

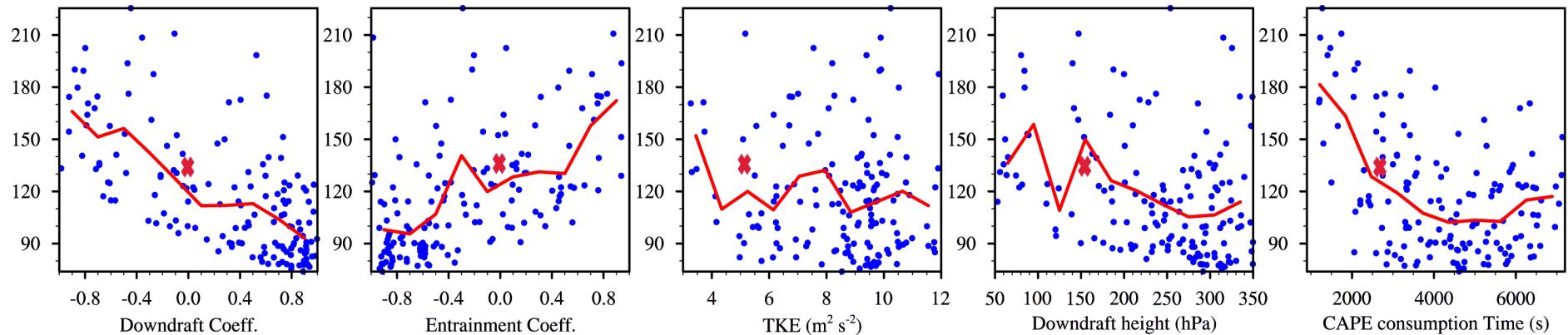
- ▶ Best model configuration
- ▶ Optimal parameters
- ▶ Sensitivity of other variables to tuned cloud parameters
- ▶ Dependence of Optimization on model grid spacing



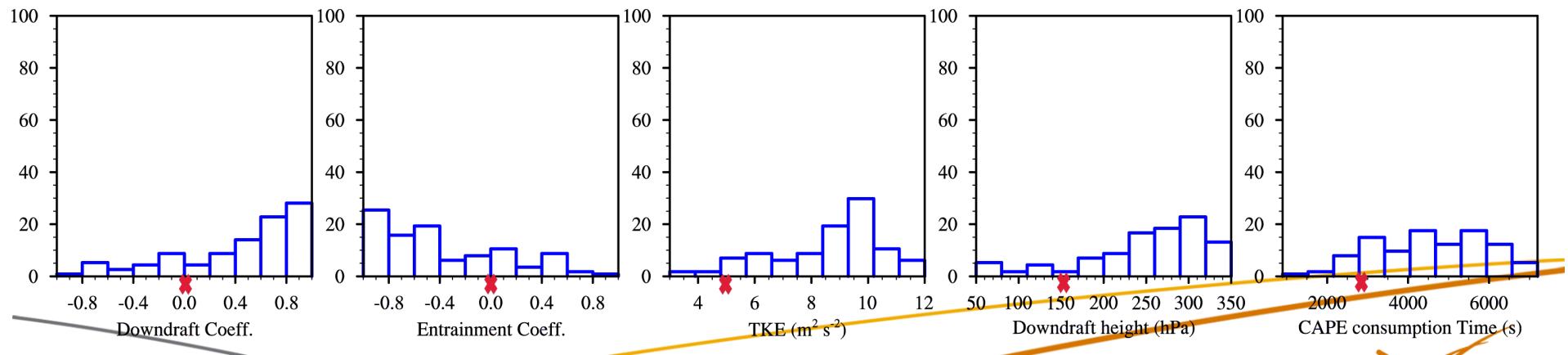
Pacific Northwest  
NATIONAL LABORATORY

# Optimal parameters

EM



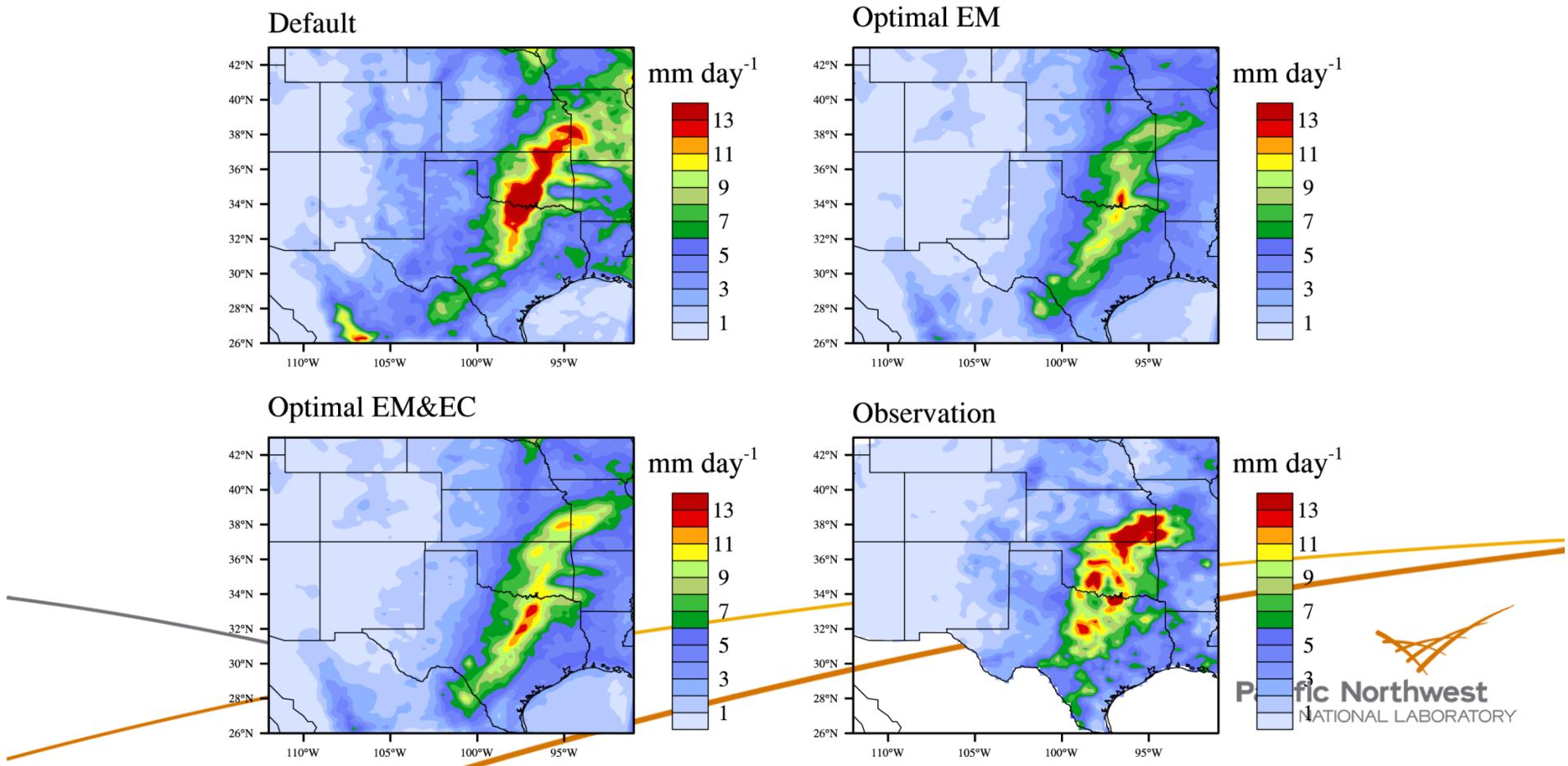
Frequency of “good” experiments



# Optimal parameters

|               | Pd   | Pe    | Pt   | Ph  | Pc   | EM  | EC   |
|---------------|------|-------|------|-----|------|-----|------|
| Default       | 0    | 0     | 5    | 150 | 2700 | 137 | 0.3  |
| Optimal EM    | 0.89 | -0.91 | 8.54 | 292 | 4615 | 74  | 0.34 |
| Optimal EM&EC | 0.57 | -0.72 | 8.9  | 321 | 3597 | 79  | 0.36 |

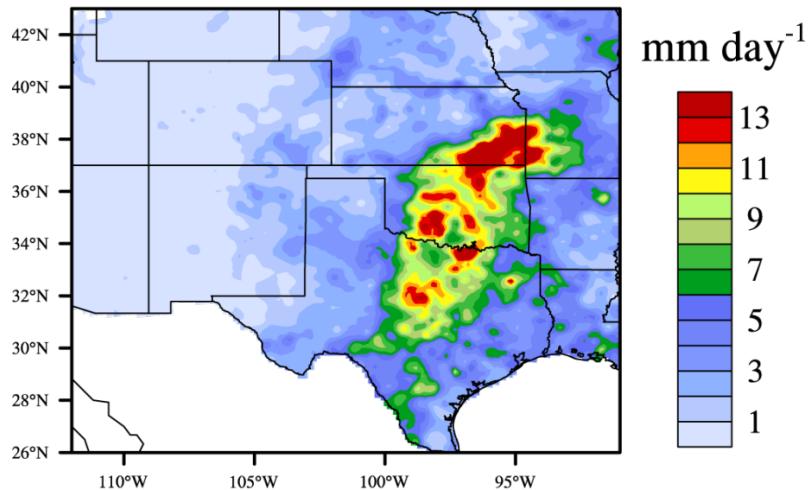
## Precipitation in June 2007



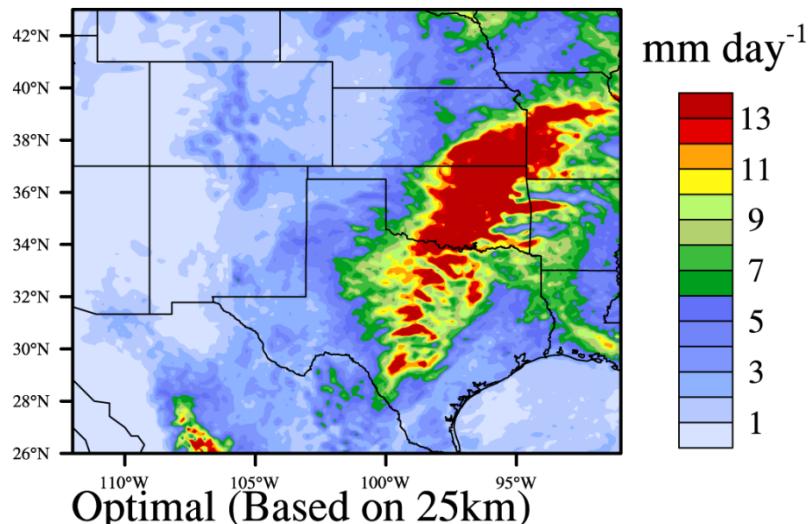
# Dependence of Optimization on model grid spacing (25km vs. 12km)

Total Precipitation in June 2007

Observation

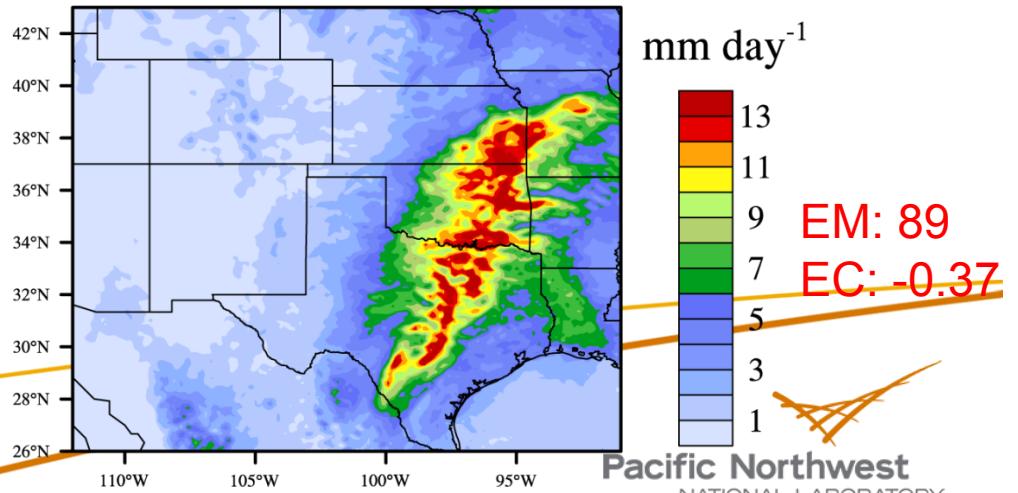


Default



EM: 148  
EC: -0.3

Optimal (Based on 25km)

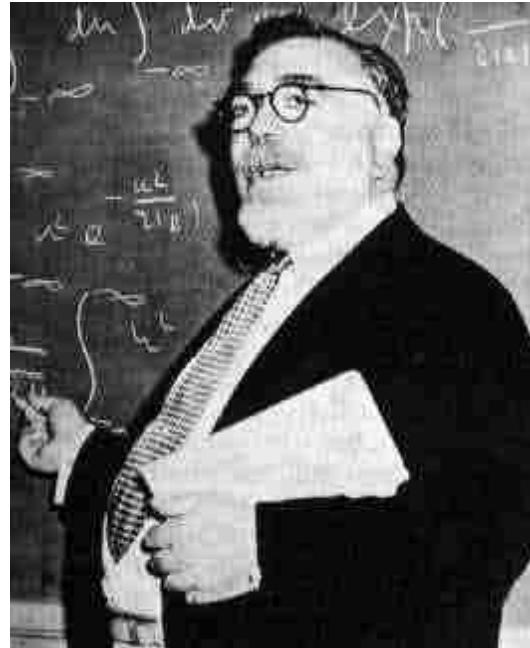


Pacific Northwest  
NATIONAL LABORATORY

## Summary - Discussion

- MEgPC & ME-PCM methods improve the convergence of gPC significantly for low regularity problems through adaptive hp-convergence.
- Sensitivity Analysis can help identify the most important parameters and suggest possible experiments.
- Sensitivity Analysis based methods, MEPCM-A & PCM-RDD improve the capability of solving high-dimensional stochastic problem.
- MEgPC, ME-PCM, MEPCM-A & PCM-RDD have been applied to solve complex high-dimensional stochastic PDE systems, such as random roughness problem & flow and transport in random porous media.
- MVFSA method can determine the optimal parameters in convective cloud scheme and qualify the sensitivity of simulated precipitation to those parameters.

# Algorithms, Analysis and Applications for High Dimensional Stochastic PDE Systems at the Exascale



*"...Because I had worked in the closest possible ways with physicists and engineers, I knew that our data can never be precise..."*

Norbert Wiener